



TECHNICKÁ UNIVERZITA V LIBERCI
Ekonomická fakulta



NÁVRH APLIKACE BUSINESS INTELLIGENCE PRO SPOLEČNOST BREX S. R. O.

Bakalářská práce

Studijní program: B6209 – Systémové inženýrství a informatika

Studijní obor: 6209R021 – Manažerská informatika

Autor práce: **Vojtěch Lank**

Vedoucí práce: Ing. Vladimíra Zádová, Ph.D.





BUSINESS INTELLIGENCE APPLICATION DESIGN FOR THE BREX S. R. O. COMPANY

Bachelor thesis

Study programme: B6209 – System Engineering and Informatics

Study branch: 6209R021 – Managerial Informatics

Author: **Vojtěch Lank**

Supervisor: Ing. Vladimíra Zádová, Ph.D.



ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Vojtěch Lank**
Osobní číslo: **E11000512**
Studijní program: **B6209 Systémové inženýrství a informatika**
Studijní obor: **Manažerská informatika**
Název tématu: **Návrh aplikace Business Intelligence pro společnost BREX s. r. o.**
Zadávací katedra: **Katedra informatiky**

Z á s a d y p r o v y p r a c o v á n í :

1. BI - charakteristika, komponenty, modely
2. Přístupy k návrhu
3. Návrh vybrané BI aplikace v prostředí firmy BREX s. r. o.
4. Zhodnocení přínosu

Rozsah grafických prací:

Rozsah pracovní zprávy: 30 normostran

Forma zpracování bakalářské práce: tištěná/elektronická

Seznam odborné literatury:

LABERGE, Robert. Datové sklady: agilní metody a business intelligence.

1. vyd. Brno: Computer Press, 2012. ISBN 978-80-251-3729-1.

POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. Business intelligence v podnikové praxi. 1. vyd. Praha: Professional Publishing, 2012.

ISBN 978-80-7431-065-2.

KIMBALL, Ralph. The data warehouse lifecycle toolkit. 2nd ed. Indianapolis, IN: Wiley Publishing, 2008. ISBN 04-701-4977-9.

Elektronická databáze článků ProQuest (knihovna.tul.cz).

Vedoucí bakalářské práce:

Ing. Vladimíra Zádová, Ph.D.

Katedra informatiky

Konzultant bakalářské práce:

Ing. Petr Bartoň

Valbek s. r. o.

Datum zadání bakalářské práce: 31. října 2013

Termín odevzdání bakalářské práce: 7. května 2014

M. Ž.

doc. Ing. Miroslav Žižka, Ph.D.
děkan



J. Skrbek

doc. Ing. Jan Skrbek, Dr.
vedoucí katedry

V Liberci dne 31. října 2013

Prohlášení

Byl jsem seznámen s tím, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu TUL.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Bakalářskou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Poděkování

Rád bych touto cestou poděkoval své vedoucí bakalářské práce paní Ing. Vladimíře Zádové, Ph.D. za cenné rady, věcné připomínky a ochotu, kterou mi v průběhu zpracování věnovala.

Dále bych rád poděkoval firmě Valbek, spol. s r.o. za umožnění práce na projektu budování datového skladu, při jehož plnění jsem získal mnoho cenných zkušeností.

Anotace

Bakalářská práce je zaměřena na problematiku Business Intelligence. Cílem práce je vytvoření návrhu Business Intelligence aplikace pro zcela specifické požadavky společnosti B R E X, s.r.o., které se vymykají všem standardně nabízeným řešením.

V teoretickém úvodu do problematiky jsou rozebrány základní komponenty Business Intelligence řešení, kterými jsou: datový sklad, datová pumpa, OLAP datové kostky, řízení kmenových dat, dolování dat a reporty.

Praktická část práce poté pojednává o návrhu konkrétních řešení jednotlivých cílů práce. Nejprve je provedena analýza požadavků, poté volba realizační platformy a následně návrh datového modelu databáze datového skladu. V praktické části se dále nachází návrh datové pumpy, OLAP datové kostky a reportů v rámci Microsoft SQL Serveru 2008 R2.

Hlavním přínosem práce je především vytvoření datového modelu specificky zaměřeného na analýzu dat ve stavebnictví, což může při budoucí realizaci navrženého Business Intelligence řešení firmě umožnit mnohem lepší a detailnější sledování nákladů a tím pádem i zvýšit konkurenceschopnost na trhu stavebních firem.

Klíčová slova

Business Intelligence, datový sklad, OLAP, datová pumpa, Kimball, multidimenzionální modelování

Annotation

This bachelor thesis focuses on the problem of Business Intelligence. The aim of the work is to create a design of Business Intelligence application for very specific requirements of B R E X, s.r.o. company, which are out of the range of commonly offered solutions.

The theoretical part of this work talks about the basic components of Business Intelligence solutions such as Data Warehouse, ETL process, OLAP Cubes, Master Data Management, Data Mining and reports.

The practical part deals with the design of particular solutions of individual goals. At first, the analysis of requirements is made, after that comes the choice of the right realisation platform and subsequently a design of data model of the Data Warehouse is presented. Later on, the practical part also suggests a design of the ETL process, OLAP Cube and reports within Microsoft SQL Server 2008 R2.

The main benefit of this work is primarily the creation of a data model which is specifically focused on data analysis in civil engineering. After this designed Business Intelligence solution is brought to life, this can allow the company to perform much better expense monitoring and therefore increase its competitiveness on the market of building companies.

Key words

Business Intelligence, Data Warehouse, OLAP, Extract Transform Load (ETL), Kimball, multidimensional modeling

Obsah

Seznam ilustrací.....	12
Seznam tabulek.....	14
Seznam použitých zkratk.....	15
Úvod.....	17
1 Zhodnocení současného stavu	18
2 Úvod do problematiky BI	20
2.1 Charakteristika BI.....	20
2.2 Komponenty BI.....	21
2.2.1 Datový sklad – DWH	22
2.2.2 Datová pumpa – ETL	23
2.2.2.1 ETL vs. ELT.....	24
2.2.3 Master Data Management.....	25
2.2.4 OLAP – datové kostky	26
2.2.4.1 Metody ukládání dat v OLAP databázích	27
2.2.4.2 Navigace v datové kostce.....	27
2.2.5 Dolování dat	28
2.2.6 Reporty	29
3 Přístupy k návrhu datových skladů.....	30
3.1 Ralph Kimball vs. Bill Inmon.....	30
3.1.1 Data Warehouse vs. Data Mart.....	30
3.2 Multidimenzionální modelování.....	31
3.2.1 Tabulky faktů.....	31
3.2.2 Typy tabulek faktů.....	32
3.2.3 Tabulky dimenzí	33
3.2.4 Hierarchie dimenzí	33
3.2.5 Typy tabulek dimenzí	34

3.3	Typy multidimenzionálních modelů	35
3.3.1	Konformní fakty a dimenze	36
3.4	Postup návrhu	36
4	Návrh řešení BI aplikace v prostředí společnosti BREX	39
4.1	Prostředí firmy	39
4.1.1	Má pozice	40
4.2	Raná studie.....	41
4.3	Analýza požadavků.....	41
4.3.1	Základní požadavky na aplikaci	41
4.3.2	Požadavky na dotazy	42
4.4	Analýza zdrojů.....	43
4.4.1	Aspe	43
4.4.2	KARAT	44
4.5	Volba realizační platformy	44
4.5.1	Microsoft SQL Server 2008 R2.....	45
4.6	Návrh datového modelu databáze datového skladu.....	46
4.6.1	Dimenze.....	47
4.6.1.1	Čas.....	48
4.6.1.2	Útvary.....	49
4.6.1.3	Stavby.....	50
4.6.1.4	Partneři	51
4.6.1.5	Dodatky	53
4.6.1.6	Kategorie ZBV	53
4.6.1.7	Zjišťovací protokoly.....	54
4.6.1.8	Scénáře	55
4.6.1.9	Potřeby	56
4.6.1.10	Výsledkové skupiny	57
4.6.2	Fakty	58
4.6.3	Fyzický model databáze	60

4.7	Návrh ETL	61
4.7.1	Plnění dimenzí	62
4.7.2	SQL Server Master Data Services	63
4.7.2.1	Plnění tabulek faktů.....	66
4.7.3	SQL Server Integration Services	66
4.8	Modelování datové kostky	68
4.8.1	SQL Server Analysis Services.....	68
4.8.2	Dimenze.....	68
4.8.3	Fakty	68
4.9	Reporty.....	70
4.9.1	SQL Server Reporting Services.....	70
5	Zhodnocení přínosu.....	72
5.1	Ekonomické zhodnocení.....	72
5.1.1	Pořizovací náklady	72
5.1.2	Provozní náklady	74
5.1.3	Návratnost investice	75
	Závěr.....	76
	Seznam použité literatury	78
	Citace.....	78
	Bibliografie.....	80
	Seznam příloh	81

Seznam ilustrací

Obrázek 2-1: Schéma fungování BI aplikace	21
Obrázek 2-2: Model fungování MDM	25
Obrázek 2-3: Znázornění datové kostky.....	27
Obrázek 3-1: DWH a DM podle Kimballa a Inmona.....	30
Obrázek 3-2: De-normalizovaná a normalizovaná dimenze	33
Obrázek 3-3: Multidimenzionální model typu hvězda (star schema).....	35
Obrázek 3-4: Multidimenzionální model typu sněhová vločka (snow-flake schema)	35
Obrázek 3-5: Multidimenzionální model typu souhvězdí (constellation schema)	35
Obrázek 3-6: Proces návrhu dimenzionálního modelu.....	38
Obrázek 4-1: Struktura společnosti	40
Obrázek 4-2: Koncept BI aplikace společnosti BREX.....	44
Obrázek 4-3: Moduly MS SQL Serveru 2008 R2	45
Obrázek 4-4: Atributy časové dimenze	48
Obrázek 4-5: Vztahy mezi atributy v časové dimenzi.....	48
Obrázek 4-6: Hierarchie v časové dimenzi.....	48
Obrázek 4-7: Atributy v dimenzi útvarů.....	49
Obrázek 4-8: Vztahy mezi atributy v dimenzi útvarů	49
Obrázek 4-9: Hierarchie v dimenzi útvarů	49
Obrázek 4-10: Atributy v dimenzi staveb	50
Obrázek 4-11: Vztahy mezi atributy v dimenzi staveb	51
Obrázek 4-12: Hierarchie v dimenzi staveb	51
Obrázek 4-13: Atributy v dimenzi partnerů	52
Obrázek 4-14: Vztahy mezi atributy v dimenzi partnerů	52
Obrázek 4-15: Hierarchie v dimenzi partnerů	52
Obrázek 4-16: Atributy v dimenzi dodatků	53
Obrázek 4-17: Vztahy mezi atributy v dimenzi dodatků.....	53
Obrázek 4-18: Atributy dimenze kategorií ZBV	54
Obrázek 4-19: Vztahy mezi atributy v dimenzi kategorií ZBV	54
Obrázek 4-20: Atributy v dimenzi zjišťovacích protokolů	54

Obrázek 4-21: Vztahy mezi atributy v dimenzi zjišťovacích protokolů	55
Obrázek 4-22: Atributy v dimenzi scénářů.....	55
Obrázek 4-23: Vztahy mezi atributy v dimenzi scénářů	55
Obrázek 4-24: Atributy v dimenzi potřeb.....	56
Obrázek 4-25: Vztahy mezi atributy v dimenzi potřeb	56
Obrázek 4-26: Hierarchie v dimenzi potřeb	57
Obrázek 4-27: Atributy v dimenzi výsledkových skupin.....	57
Obrázek 4-28: Vztahy mezi atributy v dimenzi výsledkových skupin.....	58
Obrázek 4-29: Návrh tabulky fact_stavby_cerpani	60
Obrázek 4-30: Návrh tabulek dimenzí staveb	61
Obrázek 4-31: Schéma funkce mapování záznamů.....	62
Obrázek 4-32: Návrh datového modelu MDS pro mapování výsledkových skupin.....	64
Obrázek 4-33: Návrh datového modelu MDS pro mapování útvarů.....	64
Obrázek 4-34: Návrh datového modelu MDS pro mapování zakázek na stavby.....	65
Obrázek 4-35: Návrh datového modelu MDS pro mapování obchodních partnerů.....	65
Obrázek 4-36: Návrh ETL v SSIS	67
Obrázek 4-37: Návrh kalkulovaných faktů v SSAS.....	69
Obrázek 4-38: Návrh reportu rekapitulace čerpání	71
Obrázek 4-39: Návrh reportu detailu čerpání stavby.....	71

Seznam tabulek

Tabulka 3-1: BUS matice	37
Tabulka 4-1: BUS matice podnikového datového skladu společnosti BREX.....	47
Tabulka 4-2: Detailnější popis definovaných faktů.....	58
Tabulka 4-3: Kalkulované fakty	69
Tabulka 5-1: Rozpis pořizovacích nákladů	73

Seznam použitých zkratk

Zkratka	Termín
2NF	2nd Normal Form – druhá normální forma. Pravidlo používané při normalizaci relační databáze.
3NF	3rd Normal Form – třetí normální forma. Pravidlo používané při normalizaci relační databáze určující způsob návrhu databáze pro zamezení vzniku duplicit a aktualizčních anomálií.
ASW	Aplikační Software
BI	Business Intelligence – aplikace pro podporu rozhodování.
BIDS	Business Intelligence Development Studio
BSC	Balanced Scorecard
CRM	Customer Relationship Management – aplikace pro správu vztahu se zákazníky.
DM	Data Mart – datové tržiště.
DTC	Distributed Transaction Coordinator – koordinátor distribuovaných transakcí.
DWH	Data Warehouse – datový sklad.
ELT	Extract, Load and Transform – proces přenosu dat ze zdrojových systémů do datového skladu.
ERP	Enterprise Resource Planning – aplikace pro plánování podnikových zdrojů.
ETL	Extract, Transform and Load – proces přenosu dat ze zdrojových systémů do datového skladu.
FK	Foreign Key – cizí klíč. Definuje vztah mezi dvěma databázovými tabulkami.
HOLAP	Hybrid OLAP
KPI	Key Performace Indicator – klíčový ukazatel výkonnosti.
MD	Master databáze – databáze referenčních hodnot používaná v MDM.
MDM	Master Data Management – řízení kmenových dat.
MDS	Master Data Services – nástroj pro zajištění MDM, který je součástí MS SQL Serveru 2008 R2 Enterprise.
MDX	Multi-Dimensional eXpression – dotazovací jazyk (podobný SQL) na dotazování do datových kostek v OLAP databázích.
MOLAP	Multidimensional OLAP
MS	Microsoft
OC	Odbytová cena
OLAP	Online Analytical Processing – Provádění analýzy dat z datového skladu v reálném čase.

Zkratka	Termín
OLTP	Online Transaction Processing – Počítačové zpracování transakcí v reálném čase.
PK	Primary Key – primární klíč. Slouží k jednoznačné identifikaci záznamu v databázové tabulce.
PKV	Položka kalkulačního vzorce
ROLAP	Relational OLAP
RSS	Rich Site Summary / Really Simple Syndication – webová služba pro snadné získávání zkráceného obsahu nejnovějších zpráv z webových stránek.
SCM	Supply Chain Management – aplikace pro řízení dodavatelských vztahů.
SOD	Smlouva o díle
SQL	Structured Query Language – standardizovaný dotazovací jazyk pro práci s relačními (OLTP) databázemi.
SŘBD	Systém řízení báze dat. Software zajišťující práci s databází a zprostředkovávající komunikaci mezi aplikací a uloženými daty.
SSAS	SQL Server Analysis Services – nástroj pro vytváření OLAP databází a datových kostek od společnosti Microsoft.
SSIS	SQL Server Integration Services – nástroj pro vytváření ETL od společnosti Microsoft.
SSMS	SQL Server Management Studio – nástroj pro správu databází společnosti Microsoft.
SSRS	SQL Server Reporting Services – nástroj pro vytváření reportů z různých datových zdrojů.
XML	Extensible Markup Language – datový formát pro výměnu libovolné struktury dat.
ZBV	Změny během výstavby ve stavebním projektu
ZP	Zjišťovací protokol

Úvod

Aplikace typu Business Intelligence (BI) jsou jednou z nejrychleji rostoucích oblastí podnikové informatiky. Firmám dávají větší kontrolu nad svými daty, širší a komplexnější možnosti analýz, a tím pádem značnou konkurenční výhodu. Na trhu je dostupných mnoho BI aplikací, ale většinou jsou nabízeny jako nadstavba nad konkrétními systémy. Pokud si firma žádá zcela individuální řešení, musí sáhnout po vlastním vývoji, určité formě dodavatelského řešení nebo značné modifikaci stávajícího nabízeného řešení.

Tato práce má za cíl vytvořit komplexní návrh celé aplikace BI pro stavební společnost B R E X, s. r. o., která vyžaduje zcela individuální přístup, neboť neexistuje řešení, které by její potřeby naplnilo. Jako dílčí cíle jsou stanoveny: návrh databáze datového skladu, návrh datové pumpy pro integraci výrobního aplikačního softwaru (ASW) Aspe a ERP KARAT, návrh modelu OLAP datové kostky a vytvoření vzorových reportů.

Obsah práce je členěn do teoretické a praktické části. V prvně zmíněné jsou uvedeny základní charakteristiky a komponenty BI. Praktická část se poté zabývá samotným návrhem, výběrem určitých řešení a návrhem implementace systému. V práci je zařazena i kapitola věnovaná tzv. Master Data Managementu (MDM) neboli řízení kmenových dat. Otázka MDM je na místě vždy, když se jedná o sjednocování dat z různých systémů.

Toto téma jsem si vybral zejména kvůli rostoucímu zájmu firem o BI systémy a individuální řešení datových skladů pro integraci heterogenních podnikových dat z různých provozních systémů. Dále také díky zájmu o databázové technologie, analýzu informací, hromadné zpracování a práci s velkým objemem dat.

1 Zhodnocení současného stavu

V současné době je možné dohledat velký počet publikací, závěrečných vysokoškolských prací a odborných článků na téma problematiky BI. Jedná se o velmi žádané a často diskutované téma, avšak poměrně obecné, neboť specifické požadavky se u každého průmyslového nebo obchodního odvětví velmi liší. Konkrétními aplikacemi BI ve stavebních společnostech se zabývá pouze jediná nalezená práce od jistého studenta VŠE v Praze – Jana Melichara. Jeho diplomová práce nese název „*Implementace Business Intelligence ve stavebnictví*“ a byla vypracována před šesti lety. Tato práce je především zaměřena na analýzu a definici strategických cílů stavební společnosti pomocí konceptu Balanced Scorecard (BSC) a konkrétní aplikaci BI postupů v rámci Microsoft (MS) SQL Serveru 2000.¹ Jelikož je tato práce zaměřena především na metodiku BSC a autorem vytvořený datový model byl zcela odlišný od požadavků této práce, nebylo možné se touto prací příliš inspirovat.

Při vyhledávání dalších dokumentů, zaměřených na tematiku aplikací BI ve stavebních firmách v knihovní databázi ProQuest, nebyl nalezen žádný podobný, který by zadaným požadavkům odpovídal.

Velké množství dalších publikací, které nejsou přímo zaměřeny na BI ve stavebnictví, v podstatě popisují stále stejné postupy a metodiky, které jsou z velké části převzaty od dvou nejvýznamnějších autorů, jimiž jsou Ralph Kimball a Bill Inmon. Ti jsou označováni za zakladatele datových skladů, což je nedílná součást BI systémů. Mezi nejznámější knihy, na kterých se Ralph Kimball podílel, patří již třetí aktualizované vydání z roku 2013 zvané: „*The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition*“. Tato kniha velmi důkladně popisuje metody přístupu k návrhu a budování podnikových datových skladů, a také poskytuje množství názorných příkladů a vzorových řešení. Ta se bohužel týkají pouze oblastí, do kterých stavební společnosti nespadají.

¹ MELICHAR, Jan. Implementace Business Intelligence ve stavebnictví. Praha, 2008. Dostupné z: http://www.vse.cz/vskp/show_file.php?soubor_id=23910. Diplomová práce. Vysoká škola ekonomická v Praze.

Na trhu je dnes možné nalézt velké množství hotových BI řešení. Nabízeny jsou buď samotné komplexní datové modely, nebo přímo celá řešení připravená k okamžité implementaci.² I přes svou velkou komplexnost ale nemusí nabízené řešení přesně odpovídat procesům, které jsou ve firmě již dlouhodobě zavedeny. V takovém případě se firma musí buď přizpůsobit nabízenému řešení, nebo zvolit cestu vlastního vývoje.

Z důvodu mála dostupných informací na toto konkrétní téma, jisté rizikovosti projektu a zcela specifických požadavků se zadávající firma rozhodla pro vytvoření vlastního individuálního řešení. Výhodou tohoto řešení je větší kontrola nad projektem, možnost specifikace mnohem konkrétnějších požadavků a také možné nižší pořizovací náklady, než při zpracování externí specializovanou firmou.

² LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. 1. vyd. Brno: Computer Press, 2012, s. 197. ISBN 978-80-251-3729-1.

2 Úvod do problematiky BI

Druhá kapitola teoretické části práce se zabývá základními pojmy používanými ve spojení s pojmem BI. Je zaměřena na charakteristiku základních komponent, ze kterých se takové systémy skládají.

2.1 Charakteristika BI

Pojem Business Intelligence je znám již od roku 1958, kdy jej vědecký pracovník společnosti IBM, Hans Peter Luhn, definoval takto: „*The ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal.*“³ Tedy schopnost vnímat vzájemné vztahy prezentovaných faktů takovým způsobem, aby vedly činnosti za požadovaným cílem. Smysl pojmu přetrvává dodnes, ale definici jeho dnešní podoby přinesl v roce 1989 analytik společnosti Gartner, Howard Dresner. Jeho definice zní následovně: „*A broad category of software and solutions for gathering, consolidating, analysing and providing access to data in a way that lets enterprise users make better business decisions.*“⁴ Tato definice říká, že BI je sjednocující pojem pro širokou škálu programů a řešení pro získávání, konsolidaci, analýzu a poskytování přístupu k datům tak, aby umožňovaly podnikovým uživatelům dělat lepší obchodní rozhodnutí.

V dnešní době se jedná o jednu z nejrychleji se rozvíjejících oblastí podnikové informatiky. Stále více firem přichází na to, že díky konsolidaci důležitých podnikových dat mohou mnohem efektivněji a rychleji provádět jejich analýzu, což jim může přinést značnou konkurenční výhodu a zvýšit konkurenceschopnost.

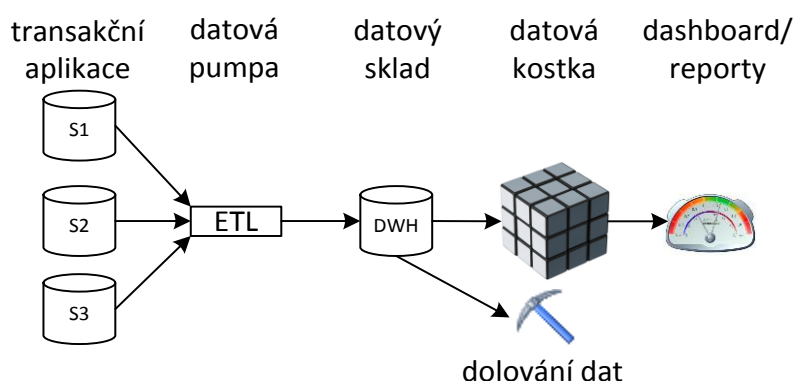
³ LUHN, H. P. A Business Intelligence System. *IBM Journal of Research and Development*. 1958, vol. 2, issue 4, s. 314-319. DOI: 10.1147/rd.24.0314. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5392644>

⁴ CHEE, Timothy. BUSINESS INTELLIGENCE SYSTEMS: STATE-OF-THE-ART REVIEW AND CONTEMPORARY APPLICATIONS. In: *Symposium on Progress in Information & Communication Technology 2009*. Malaysia: xxx, 2009, s. 96. Dostupné z: http://spict.utar.edu.my/SPICT-09CD/contents/pdf/SPICT09_A-5_1.pdf

Aplikace BI slouží nejen pro podporu rozhodování vrcholových pracovníků, ale i pro sledování vývoje klíčových ukazatelů výkonnosti (KPI). Primárním úkolem těchto aplikací je nabídnout uživateli rychlý přehled o aktuálním stavu fungování podniku, získání detailních reportů podle zadaných kritérií nebo dostat včasné varování, pokud je zaznamenána změna oproti plánu.

2.2 Komponenty BI

Dnešní aplikace BI bývají velmi důmyslnými a komplexními nástroji pro analýzu a interpretaci velkého množství podnikových dat. Takové aplikace se skládají z několika dílčích systémů, z nichž každý má ve výsledném produktu své nezastupitelné místo. Základními stavebními kameny jsou: datový sklad – Data Warehouse (DWH), datová pumpa (ETL), analytické databáze – On Line Analytical Processing (OLAP), dolování dat a reporty. Obrázek 2-1 schematicky znázorňuje princip fungování aplikací typu BI.



Obrázek 2-1: Schéma fungování BI aplikace

Zdroj: vlastní

Data jsou nejprve pomocí ETL načtena z různých zdrojů do DWH, který slouží jako zdroj dat pro OLAP databáze, které dále slouží jako zdroj dat pro informativní dashboardy a reporty. Data umístěná v DWH mohou také sloužit jako zdroj dat pro aplikace umožňující dolování dat.

Teoretická část je dále nejprve zaměřena na DWH, poté na jeho plnění pomocí ETL procesů, OLAP analýzu, MDM, dolování dat a reporty.

2.2.1 Datový sklad – DWH

V běžné praxi mají firmy hned několik systémů pro správu a uchovávání dat důležitých pro rozhodování a plánování. Například Enterprise Resource Planning – ERP (aplikace pro plánování podnikových zdrojů), Customer Relationship Management – CRM (aplikace pro správu vztahu se zákazníky), Supply Chain Management – SCM (aplikace pro řízení dodavatelských vztahů) a mnoho dalších. Aby bylo možné tato podniková data hromadně analyzovat, je potřeba, je dostat ze všech provozních systému na jediné místo. Takové místo se nazývá datový sklad. Jedná se o specializovanou databázi pro shromažďování dat z různých podnikových systémů v takové formě, aby nad nimi bylo možné rychle a snadno provádět analytické dotazy.

Existuje mnoho definic datového skladu, ale asi nejznámější a nejvýstižnější je definice jednoho ze zakladatelů data warehousingu – Billa Inmona. DWH charakterizuje jako „*subjektově orientovanou, integrovanou, stálou a časově rozdílnou kolekci dat pro podporu rozhodování manažerů.*“⁵ Subjektovou orientací se rozumí orientace na takový subjekt, podle kterého jsou data v DWH kategorizována. Tím může být zákazník, zaměstnanec, výrobek, územní jednotka a podobně.⁶

Databáze DWH vychází z relačního modelu dat, ale tento termín se zde nepoužívá, hovoří se o multidimenzionálním modelu dat, multidimenzionálním schématu, multidimenzionální databázi. Multidimenzionální schéma je tvořeno tabulkami dimenzí a tabulkami faktů. Oba typy tabulek jsou databázové relace s určitými specifiky, které zohledňují cíl, pro který jsou určeny. Mohou vytvářet hvězdicové struktury (star schema), různé formy sněhových vloček (snowflake schema) a souhvězdí (constellation schema). Podrobnější rozbor multidimenzionálního modelování se nachází v kapitole 3.2.

Zatímco v OLTP databázích je kladen důraz především na rychlé operace s jednotlivými řádky (výběr, vložení, úprava, odstranění), u DWH jsou data vkládána nebo aktualizována

⁵ ZÁDOVÁ, Vladimíra. *Specifika postavení a návrhu datových skladů v rámci IS/ICT*. Liberec, 2006. Disertační. Technická univerzita v Liberci, s. 24.

⁶ HORÁK, Jiří a Bronislava HORÁKOVÁ. *Datové sklady a využití datové struktury typu hvězda pro prostorová data*. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, 2007. Dostupné z: http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2007/sbornik/Referaty/Sekce3/hvezdaF4.pdf

(upravována) pouze jednou za určitou periodu, a to ve velkých objemech. Jelikož poté v drtivé většině případů dochází pouze ke čtení dat, je mnohem větší důraz kladen právě na rychlost získávání dat (čtení).

DWH může být realizován na stejných platformách jako transakční databáze (MS SQL Server, Oracle atd.) nebo pomocí speciálně navržených systémů pro řízení báze dat (SŘBD) určených přímo pro realizaci datových skladů. Takový produkt nabízí například společnost Teradata.⁷

V DWH nejsou ukládána data pouze aktuální, ale především historická. Do DWH se informace dostávají zpravidla dávkově, v přesně stanovených časových intervalech a zpravidla dochází k jejich přidávání nebo změně. Odstraňování záznamů se zde příliš nevyužívá, z důvodu zachování historických informací pro možnost zpětného sledování změn. Samozřejmě za předpokladu, že je zachování historických údajů žádoucí.

2.2.2 Datová pumpa – ETL

Pro přenos dat mezi provozními systémy a DWH se používají specializované nástroje zvané datové pumpy. Tento výraz je volným překladem anglického „*Extract, Transform and Load*“ (ETL). Tento nástroj data nejprve načte, poté transformuje a nakonec uloží do DWH.

Jedná se o specifické aplikace navržené tak, aby uměly načíst předem definovanou strukturu dat z různých zdrojů a po patřičných úpravách, očištění dat a standardizaci nebo doplnění chybějících hodnot, tato data uložily do DWH. Jako zdroje dat mohou posloužit OLTP databáze, XML soubory, soubory tabulkových procesorů (Excel, Calc), textové soubory tzv. „flat files“ (txt, csv) nebo dokonce webové stránky či RSS kanály.

Na tuto komponentu BI řešení bývá z pravidla kladen největší důraz, neboť na kvalitě „očištění“ dat závisí i kvalita veškerých výstupů, analýz a reportů. Pokud se tedy do DWH dostanou špatně interpretovaná nebo dokonce nesprávná data, může to mít značně

⁷ TERADATA. Teradata Database: Teradata Database 15.0. *TERADATA* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.teradata.com/products-and-services/Teradata-Database>

negativní dopad na rozhodovací možnosti řídicích pracovníků. Tím, že je ETL klíčovým pro správné fungování celého BI systému, zabírá jeho vývoj nejvíce času z celého projektu. Mluví se o 60 – 80 %⁸.

Důležité je zmínit, že úkolem ETL je vybrat ze zdrojových dat jen tu část, která je potřeba pro uložení v DWH. V žádném případě není žádoucí načíst všechna data a uložit pouze jejich zlomek. Nesprávná selekce může celý proces značně zpomalit nebo dokonce zahltit zdrojové databáze, a tím pádem znemožnit běh aplikací, které s nimi pracují.

Klasický model integrace dat (ETL) využívá buď transformací přímo během přenosu ze zdroje do DWH nebo využívá dočasné úložiště (staging area). V takovém případě jsou data nejprve uložena do dočasné databáze, poté jsou nad nimi provedeny transformační a čistící operace a následně jsou přenesena do datového skladu.

2.2.2.1 ETL vs. ELT

V posledních letech se vedle ETL pro ukládání dat ze zdrojů do DWH používá i ELT. Jedná se o variantu datové pumpy, která nejprve načte zdrojová data do DWH a teprve poté nad nimi provádí transformační operace. Výhodou tohoto řešení je snížení času potřebného pro přenos dat mezi zdrojem a DWH a možný vyšší výkon během transformací, neboť veškeré tyto operace jsou prováděny na úrovni SŘBD. Naopak nevýhodou je mnohem nižší flexibilita celého procesu, neboť je programátor omezen možnostmi manipulovat s daty mimo databázi. ELT také na rozdíl od ETL vyžaduje dočasné tabulky, bez kterých by většina transformací nebyla možných. Asi nejzávažnějším neduhem modelu ELT je, že pokud jsou dotazy do DWH směřovány velmi často a v krátkých intervalech, může se stát, že by byl report sestaven z „neočištěných“ dat, což je samozřejmě nežádoucí. V takovém případě je vhodnější použít model ETL, který tento problém spolehlivě řeší.⁹

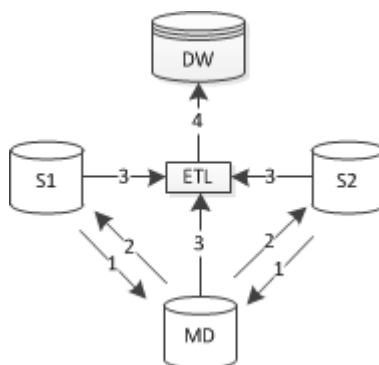
⁸ POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. *Business intelligence v podnikové praxi*. 1. vyd. Praha: Professional Publishing, 2012, s. 24. ISBN 978-80-7431-065-2.

⁹ LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. 1. vyd. Brno: Computer Press, 2012, s. 247-249. ISBN 978-80-251-3729-1.

2.2.3 Master Data Management

V případech, kdy jsou v DWH slučována data z různých provozních systémů, je na místě otázka tzv. Master Data Managementu (MDM) neboli řízení kmenových dat. MDM se využívá především v situacích, kdy je například v podniku zavedeno několik informačních systémů, kde každý nějakým způsobem pracuje například se seznamem zákazníků. Problém nastává tehdy, kdy je nový zákazník zaveden do jednoho systému a v jiných není. V takovém případě dochází velmi snadno ke vzniku duplicit a mylných informací.¹⁰

Aplikace, které umožňují MDM, mají za úkol zajistit synchronizaci seznamů (číselníků) využívaných systémy napříč celým podnikem mezi všemi těmito provozními systémy. Výhody takového řešení jsou zjevné. V podniku je vedena pouze jedna verze pravdy, data jsou centrálně uložena v master databázi (MD), čímž je mnohem jednodušší jejich správa a každá změna v kmenových datech se projeví ve všech ostatních systémech.



Obrázek 2-2: Model fungování MDM

Zdroj: vlastní

Obrázek 2-2 graficky znázorňuje model fungování MDM. V prvním kroku jsou do systému pro správu kmenových dat nahrána data z provozních systémů (S1 a S2). Poté v systému probíhá čištění a validace, po které jsou ověřené údaje zpětně zapsány do všech provozních systémů (2), čímž je zajištěna stejná kvalita dat napříč systémy. Posledním krokem (3 a 4) je sehrání dat z provozních systémů a MD do DWH.

¹⁰ POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. *Business intelligence v podnikové praxi*. 1. vyd. Praha: Professional Publishing, 2012, s. 153-158. ISBN 978-80-7431-065-2.

Díky referenčním hodnotám v MD není nutné sdílené seznamy stahovat z provozních systémů, ale jako zdroj pro dimenze poslouží přímo seznamy v MD.¹¹

Implementace MDM bývá ze začátku poměrně složitá, ale efektivita a řád, který vnese do podnikových dat, mnohonásobně převyšuje vstupní investice.

2.2.4 OLAP – datové kostky

V poslední dekádě 20. století se s rychlým nástupem informačních technologií začalo v podnicích prudce zvyšovat množství ukládaných informací, čímž se začaly zvyšovat i nároky na jejich zpracování. I nejvýkonnější servery té doby nebyly dostatečně výkonné na to, aby v rozumném čase vrátily výsledky požadavku na sumarizovaná data z transakčních databází. To mělo za následek velmi omezené možnosti analýzy dat a tím výrazně sníženou rozhodovací schopnost řídících pracovníků. Začalo se tedy přemýšlet nad tím, jak data ukládat, aby bylo možné rychle získávat agregované hodnoty podle určitých kritérií.

Na tyto požadavky reagovaly firmy (Microsoft, IBM, Oracle) různým způsobem a v roce 1998 Microsoft představil svou první komerční databázi podporující technologii OLAP. Jedná se o speciální typ databáze, která umožňuje ukládat předpočítaná agregovaná data v různých úrovních podrobnosti (granularity). Díky OLAP databázím již tedy nebylo nutné sumarizované výstupy zdlouhavě počítat při každém požadavku, čímž se výrazně snížila doba na získání takového výstupu.

V současné době je na trhu několik firem nabízejících platformy pro OLAP databáze. Většina je komerční, ale najdou se mezi nimi i volně šiřitelné verze.

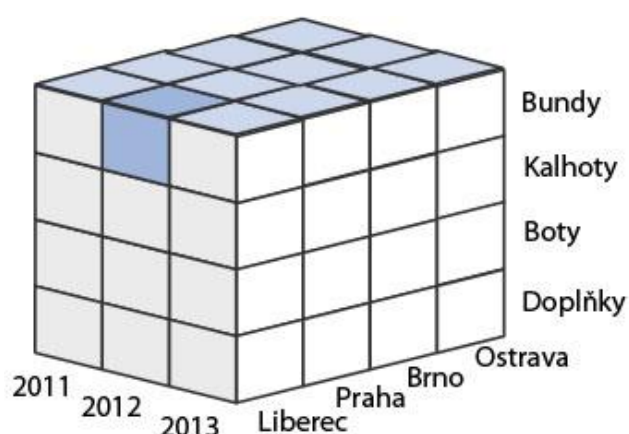
¹¹ REEVES, Laura L. *A Manager's Guide to Data Warehousing*. Indianapolis, IN: Wiley Pub., c2009, s. 240-243. ISBN 0470176385.

2.2.4.1 Metody ukládání dat v OLAP databázích

OLAP databáze využívá pro ukládání dat několik různých technik, které se liší především z hlediska použití. První z nich je MOLAP – Multidimensional OLAP. V tomto režimu jsou do binární OLAP databáze ukládána všechna data. Jak na nejnižších úrovních, tak ta agregovaná. Tato metoda je vhodná pouze do určité velikosti databáze, neboť při velkém objemu dat může výpočet datové kostky trvat neúnosně dlouhou dobu. Ovšem velkou výhodou je velmi rychlá odezva. Pro extrémně velké databáze je vhodnější použít režim ROLAP – Relational OLAP. V tomto módu pracuje OLAP databáze pouze se zdrojovými daty v DWH a veškeré agregované hodnoty jsou počítány při každém dotazu. Výhodou je využití možností OLAPu na velkých datech, ale nevýhodou velmi pomalá odezva. Posledním typem je HOLAP. Tzv. hybridní OLAP využívá kombinace obou zmíněných technologií. Agregované hodnoty jsou předpočítány v binární OLAP databázi, ale dotazy na prvky nejnižších úrovní hierarchií jsou směřovány do DWH. To umožňuje rozumný kompromis mezi rychlostí a objemem OLAP databáze.¹²

2.2.4.2 Navigace v datové kostce

OLAP databáze jsou známy především pod pojmem datové kostky (OLAP Cubes), neboť ukládají kombinace dimenzí, metrik a hierarchií do struktury znázorňované jako krychle – viz obrázek 2-3. Každý průsečík všech dimenzí znamená jednu konkrétní hodnotu.



Obrázek 2-3: Znázornění datové kostky

Zdroj: vlastní

¹² MICROSOFT. Partition Storage Modes and Processing. MICROSOFT. *Microsoft Developer Network* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://msdn.microsoft.com/en-us/library/ms174915.aspx>

Pro získávání dat z datových kostek slouží několik obecných technik.

- Slicing (krájení – omezení jedné dimenze)
- Dicing (dělení na menší kostky – omezení prvky několika dimenzí)
- Drill-down (posun v hierarchii o úroveň níže – detailnější informace)¹³
- Drill-up (posun v hierarchii o úroveň výše – vyšší agregace)¹⁴
- Drill-across (zobrazení metrik sdílející stejné dimenze)¹⁵
- Drill-through (doslova „provrtání“ kostky a zobrazení dat uložených přímo v datovém skladu, ze kterých je kostka sestavena)
- Roll-up (sumarizace fyzických nebo vypočtených faktorů podle určité dimenze)
- Pivoting (otočení kostky – změna pohledu)

2.2.5 Dolování dat

Především u obchodních společností je klíčovým bodem podnikání znalost svých zákazníků. Pomocí standardní analýzy podnikových dat je možné získat ucelené informace o stavu fungování podniku, jednotlivých poboček, nebo například podílů prodeje dle demografického členění. Avšak v těchto těžce nasbíraných datech se mnohdy ukrývají skryté souvislosti, které se bez pomoci specializovaných aplikací velmi těžce odhalují. Právě tyto souvislosti jsou pro firmu velmi cennou komoditou a mohou firmě přinést značnou konkurenční výhodu. Obor, který se hledáním skrytých souvislostí ve velkých objemech dat zabývá, se nazývá Data Mining neboli dolování dat a je definován následujícím způsobem:

*Data mining je proces výběru, prohledávání a modelování ve velkých objemech dat, sloužící k odhalení dříve neznámých vztahů mezi daty za účelem získání obchodní výhody.*¹⁶

¹³ KIMBALL, Ralph. Drilling Down, Up, and Across: Understanding the vocabulary of navigating dimensions. *KIMBALL GROUP* [online]. 1996 [cit. 2014-05-05]. Dostupné z: <http://www.kimballgroup.com/1996/03/01/drilling-down-up-and-across/>

¹⁴ Tentýž

¹⁵ Tentýž

¹⁶ DANEL, Roman. *Dolování dat* [online]. 2010, 3 s. [cit. 2014-05-05]. Dostupné z: http://home1.vsb.cz/~dan11/is_skripta/IS%202010%20-%20Danel%20-%20Dolovani%20dat.pdf

Pro dolování dat se používá různých technik vycházejících především ze statistické analýzy. Mezi základní používané metody se řadí rozhodovací stromy, neuronové sítě a regresní, shluková nebo asociační analýza. V rámci BI aplikací slouží jako zdroje dat pro data miningové algoritmy dobře strukturovaná data z DWH.¹⁷

2.2.6 Reporty

Tato podkapitola se zabývá tzv. reporty, které jsou nedílnou součástí každé BI aplikace.

Díky konsolidaci velké množiny podnikových dat nabízejí BI aplikace širokou škálu nástrojů jak pro důkladnou analýzu dat, tak pro rychlé získání KPI. Úkolem reportovacích služeb je jakýmkoliv způsobem uživateli zobrazit (tabulkově nebo graficky) „nasbíraná“ agregovaná data a na požádání zobrazit více podrobností nebo jiný úhel pohledu. Pomocí dnešních technologií je možné reporty zcela automatizovat. V takovém případě jsou automaticky generované reporty v definovaných intervalech odesílány například na email řídících pracovníků nebo publikovány na firemní intranet.

Do kategorie reportů můžeme zařadit i tzv. dashboardy (přístrojová nebo palubní deska). Jedná se o speciální kategorii reportů, které bývají realizovány pomocí interaktivních webových aplikací. Na malé ploše graficky zobrazují pouze nejdůležitější indikátory ukazující, jaký je aktuální stav firmy. Mezi typické informace, které se na dashboardech nachází, jsou například: aktuální plnění měsíčního/ročního plánu, denní prodeje, cash-flow nebo nárůst/pokles zisku oproti stejnému období v předešlém roce. Tyto hodnoty bývají nejčastěji zobrazovány pomocí budíků a grafů.

¹⁷ WITHEE, Ken. *Microsoft Business Intelligence For Dummies*. Hoboken, NJ: Wiley Pub., 2010, s. 126-128. ISBN 978-0-470-52693-4.

3 Přístupy k návrhu datových skladů

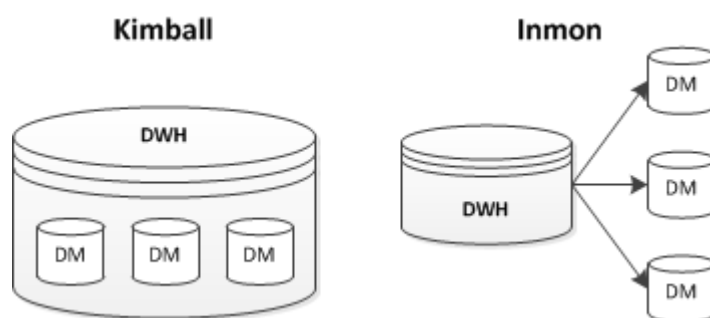
V této kapitole se nachází rozbor jednotlivých přístupů k návrhu datových skladů, jejich zakladatelé a typy používaných architektur.

3.1 Ralph Kimball vs. Bill Inmon

Ralph Kimball a Bill Inmon jsou považováni za zakladatele data warehousingu. Přestože oba mluví o DWH, každý si pod ním představuje něco trochu jiného a z těchto odlišností vycházejí i rozdílné metodologie a přístupy k návrhu modelu.

3.1.1 Data Warehouse vs. Data Mart

V souvislosti s datovými sklady se setkáváme s pojmem „datové tržiště“ neboli Data Mart (DM). A právě v chápání rozdílu mezi DWH a DM tkví neustálý boj mezi zastánci teorií Ralpha Kimballa a Billa Inmona. Obrázek 3-1 tento rozdíl znázorňuje.



Obrázek 3-1: DWH a DM podle Kimballa a Inmona

Zdroj: vlastní

„Data warehouse is the conglomerate of all data marts within the enterprise. Information is always stored in the dimensional model“¹⁸ Tak zní Kimballova teorie o datových

¹⁸ 1KEYDATA. Bill Inmon vs. Ralph Kimball. *1KeyData* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.1keydata.com/datawarehousing/inmon-kimball.html>

skladech a tržištích. Říká, že datový sklad je konglomerát všech datových tržišť celého podniku a data jsou v něm uložena ve formě (multi)dimenzionálního modelu.

Teorie Billa Inmona zní následovně: „*Data warehouse is one part of the overall business intelligence system. An enterprise has one data warehouse, and data marts source their information from the data warehouse. In the data warehouse, information is stored in 3rd normal form.*“¹⁹ Na rozdíl od předešlé definice Inmon říká, že v datovém skladu jsou data uložena ve třetí normální formě (3NF) a z něho všechna datová tržiště čerpají.

Dále se budu zabývat pouze teorií, kterou propaguje Ralph Kimball, neboť jsem jeho metodu přístupu k návrhu budování podnikových datových skladů zvolil pro vypracování praktické části této práce.

3.2 Multidimenzionální modelování

Jak bylo zmíněno v definici datového skladu, pro ukládání dat se využívá tzv. multidimenzionální model. Jedná se o speciální techniku datového modelování založenou na dvou základních typech entit – dimenzích a faktech. V následujících kapitolách jsou popsány základní principy multidimenzionálního modelování.

3.2.1 Tabulky faktů

Tabulky faktů představují v multidimenzionálním modelu entity obsahující nějaký sledovaný údaj nebo jev. Jedná se z pravidla o ekonomický ukazatel. Aby bylo možné takový údaj měřit a vyhodnocovat, musí být numerický. Jako příklad faktu může být počet prodaných kusů nebo prodejní cena. Fakty jsou rozdělovány do tří skupin. Pokud má u vyhodnocování smysl tyto konkrétně uvedené údaje sčítat za všechny dimenze (při operaci roll-up), jedná se o tzv. zcela aditivní fakty. Pokud fakt není aditivní vzhledem alespoň k jedné dimenzi, jedná se o semi-aditivní fakt. Jestliže sčítání postrádá smysl za všechny dimenze, tento fakt je označován za neaditivní.

¹⁹ 1KEYDATA. Bill Inmon vs. Ralph Kimball. *1KeyData* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.1keydata.com/datawarehousing/inmon-kimball.html>

Podle Kimballa jsou fakty neklíčové atributy v tabulce faktů. Vedle nich jsou v tabulkách faktů také klíčové atributy sloužící jako primární klíč (PK). V tabulkách faktů bývá zpravidla PK tvořen všemi atributy sloužících jako cizí klíč (FK) odkazující na PK tabulky dimenzí. Existují ale i výjimky kde PK tvoří samostatný atribut.

Za zmínku také stojí tabulky faktů neobsahující žádný fakt (factless fact table). Takové tabulky neobsahují žádný numerický údaj a v analýzách se počítá pouze s počty zaznamenaných řádků.²⁰

3.2.2 Typy tabulek faktů

Vzhledem ke granularitě faktů existují v datovém skladu tři typy tabulek faktů. Prvním typem je transakční tabulka faktů. V takové tabulce jsou shromažďovány údaje o konkrétních transakcích na nejvyšší možné úrovni podrobnosti. Jedná se například o uskutečněné prodeje nebo přístupy na webové stránky. Každý záznam se váže na jeden konkrétní časový okamžik.

Pokud není potřeba sledovat časovou osu tak podrobně, pak je tu druhý typ – periodické snímky. V takové tabulce jsou uchovávány kumulované informace za určité časové období; např. den nebo měsíc. Tento typ faktových tabulek je v praxi nejpoužívanější a velmi dobře slouží pro předpověď vývojových trendů sledovaných jevů.

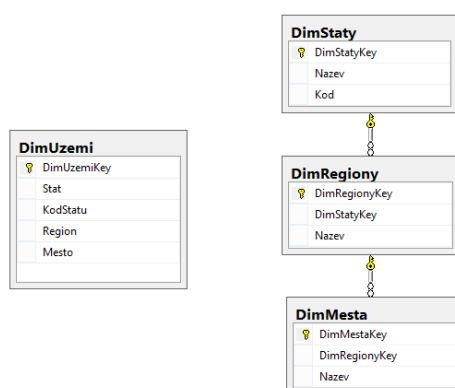
Posledním typem jsou tzv. akumulované (stavové) snímkové tabulky. Tento typ je také závislý na transakcích, ale z jiného úhlu pohledu. Do této tabulky spadají informace spojené s určitými milníky v nějakém procesu. Například postupy v procesu objednávky. Od objednání, přes odeslání zboží až po doručení.²¹

²⁰ POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. *Business intelligence v podnikové praxi*. 1. vyd. Praha: Professional Publishing, 2012, s. 74. ISBN 978-80-7431-065-2.

²¹ Tentýž s. 72.

3.2.3 Tabulky dimenzí

V tabulkách označovaných jako „dimenze“ jsou uchovávány informace, které dávají konkrétnímu faktu jeho význam. Jsou to ony subjekty podnikání, o kterých se zmiňuje Inmon v definici DWH. Pomocí atributů dimenzí, lze analyzovat výkon, tj. konkrétní ekonomický ukazatel uložený v tabulce faktů. Nejběžněji používané dimenze obsahují například údaje o čase, zákaznících nebo produktech.



Obrázek 3-2: De-normalizovaná a normalizovaná dimenze

Zdroj: vlastní

Důležité je zmínit, že tabulky dimenzí nemusejí dodržovat 3NF. Takové tabulky se nazývají de-normalizované a musejí pak dodržovat druhou normální formu (2NF). Rozdíl mezi těmito dvěma strukturami je znázorněn na obrázku výše. V levé polovině je de-normalizovaná tabulka dimenze Území a v pravé pak stejná dimenze dodržující 3NF. De-normalizované dimenze je vhodné použít tam, kde obsahují malý počet atributů. S rostoucím počtem atributů je vhodnější dimenzi normalizovat, neboť díky redundanci by mohla její velikost značně narůst. Na druhé straně, v porovnání s objemy tabulek faktů, tvoří většinou dimenze pouze zlomek celkového objemu dat v DWH, tudíž je na zvážení, zda je normalizace významná, či nikoliv.

3.2.4 Hierarchie dimenzí

Každá dimenzionální tabulka obsahuje různé atributy, kterých může být v běžné praxi až několik desítek. Z některých atributů je možné v rámci dimenze definovat tzv. hierarchie, které slouží pro vytváření agregací. Dimenzionální atributy se dělí na dva druhy. Ty, které

určují agregační úrovně hierarchií (úrovňové atributy) a ty, které pouze blíže specifikují danou úroveň v hierarchii (popisné atributy). Úrovňové atributy tedy definují konkrétní úrovně hierarchií, na kterých jsou agregace vytvářeny.²² Pomocí těchto atributů je možné provádět operace *slicing/dicing*, *drill-down/drill-up* atd. Pomocí popisných atributů tyto operace provádět nelze, neboť slouží pouze jako doplňující informace k dané úrovni.

3.2.5 Typy tabulek dimenzí

Oproti výše zmíněným normalizovaným dimenzím dodržující 3NF a de-normalizovaným dimenzím, existují i další typy. Jedním z nich je „parent-child“ dimenze. V takové tabulce existuje atribut stejného významu jako PK odkazující na PK téže tabulky, pro vytvoření nepravidelné hierarchie. Tím je možné definovat hierarchickou strukturu například firemních útvarů nebo zaměstnanců.

Dalším typem je tzv. degenerovaná dimenze. Ta obsahuje pouze jeden atribut a ten je umístěn přímo v tabulce faktů. Tento typ dimenze nachází největší uplatnění jako například dimenze obsahující číslo faktury nebo transakce.

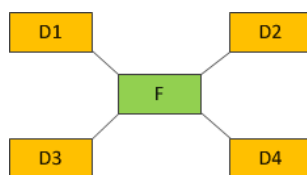
Posledním typem, který bude zmíněn, je tzv. sběrná dimenze neboli „Junk Dimension“. Některé fakty bývají často rozlišovány pomocí různých příznaků nebo kódů, například M/Ž, nebo ANO/NE. Každý z těchto příznaků v podstatě tvoří samostatnou miniaturní dimenzi. Jelikož nemají žádné další atributy a samy zabírají pouze několik znaků, vytváření mnoha takovýchto malých dimenzí nebývá příliš efektivní. Z toho důvodu je výhodnější sloučit tyto příznaky do jedné sběrné dimenze. V tabulce takové dimenze se pak nacházejí všechny možné kombinace těchto příznaků a do tabulky faktů odkazuje umělým klíčem označující unikátní kombinací těchto atributů.²³

²² ZÁDOVÁ, Vladimíra. *Specifika postavení a návrhu datových skladů v rámci IS/ICT*. Liberec, 2006. Disertační. Technická univerzita v Liberci. s. 88

²³ LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. 1. vyd. Brno: Computer Press, 2012, s. 170-179. ISBN 978-80-251-3729-1.

3.3 Typy multidimenzionálních modelů

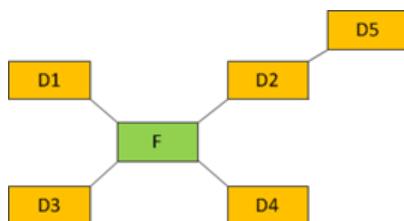
Multidimenzionální modely mohou nabývat tří různých typů. Jsou jimi hvězda, sněhová vločka a souhvězdí. Nejzákladnějším je model hvězdy neboli „star schema“. V tomto případě je v modelu jedna tabulka faktů, kterou obklopuje několik de-normalizovaných dimenzí. Následující obrázek tento model znázorňuje.



Obrázek 3-3: Multidimenzionální model typu hvězda (star schema)

Zdroj: vlastní

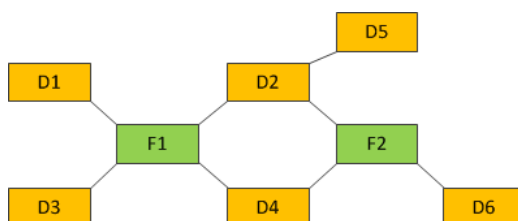
Pokud by došlo k normalizaci alespoň některé z dimenzí, nejednalo by se už o schéma hvězdy, ale nýbrž o sněhovou vločku (snow-flake schema) – viz obrázek 3-4.



Obrázek 3-4: Multidimenzionální model typu sněhová vločka (snow-flake schema)

Zdroj: vlastní

Posledním typem je souhvězdí neboli konstelace (constellation schema). Takový model obsahuje několik tabulek faktů, které mezi sebou sdílí některé dimenze. Pokud by mezi sebou sdílely všechny dimenze, dalo by se schéma zjednodušit do jednoho z předešlých. Obrázek 3-5 tento model znázorňuje.



Obrázek 3-5: Multidimenzionální model typu souhvězdí (constellation schema)

Zdroj: vlastní

Některé zdroje uvádějí²⁴, že modely typu souhvězdí obsahují pouze de-normalizované tabulky dimenzí a pokud u některých z nich dojde k normalizaci, nejedná se dále o souhvězdí, nýbrž sněhovou bouři neboli „snowstorm“.

3.3.1 Konformní fakty a dimenze

Pokud jsou v multidimenzionálním modelu některé dimenze sdílené mezi více tabulkami faktů, nazývají se konformní dimenze. Aby splňovaly požadavky konformnosti, jsou na ně kladeny speciální nároky. Taková dimenze musí být na nejnižší možné úrovni podrobnosti (granularity), aby vyhovovala všem faktům.

Podrobnost neboli granularita takové dimenze musí být rovna nejvyšší granularitě zaznamenávaného faktu. Ostatní fakty s nižší podrobností budou pouze agregací stejné dimenze. Konformní dimenze jsou tedy buď identické, nebo se jedná o striktní matematickou podmnožinu z nejvyššího detailu dimenze.²⁵ Pokud je v multidimenzionálním modelu nějaký fakt definován stejnou množinou dimenzí jako jiný a oba jsou na stejné úrovni granularity, tyto fakty jsou označovány jako konformní.²⁶

3.4 Postup návrhu

Proces návrhu dimenzionálního modelu probíhá ve čtyřech krocích. Prvním je výběr podnikových procesů, u kterých se dá BI maximálně využít, neboli stanovení rámce. To spočívá v určení oblasti dat, která se budou v DWH shromažďovat a určení dimenzí, podle kterých budou moci být data analyzována. Pro tuto analýzu se používá tzv. „Enterprise Data Warehouse BUS Matrix“ neboli sběrníková matice podnikového datového skladu, což je Kimballova metoda využívána při multidimenzionálním modelování právě pro analýzu podnikových procesů a dimenzí. Jedná se o tabulku, která v řádcích definuje podnikové procesy (fakty) a ve sloupcích použité dimenze. Každý řádek tak ve výsledném modelu tvoří jednotlivé DM. Průsečík procesu a dimenze vyznačený

²⁴ FROFINIT. *Dimenzionální modelování*. 2013. Dostupné z: http://www.profinet.eu/fileadmin/Content/profinet.eu/Academy/MI-DSP/08_Dimenzionalni_modelovani.pdf

²⁵ KIMBALL, Ralph a Margy ROSS. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Third edition. Indianapolis: John Wiley & Sons, Inc., 2013, s. 130-138. ISBN 11-185-3080-2.

²⁶ Tentýž, s. 138-139.

křížkem znázorňuje vzájemnou provázanost. Z takové matice lze jednoduše vyčíst všechny použité dimenze, konformní dimenze a konformní fakty, což slouží jako základ pro návrh logického a fyzického modelu databáze datového skladu.

Tabulka 3-1 ukazuje jednoduchou BUS matici datového skladu. Z návrhu je patrné, že konformními dimenzemi jsou „Čas“, „Produkt“ a „Obchod“ neboť jsou sdílené mezi více procesy (fakty).

Tabulka 3-1: BUS matice

	DIMENZE				
PODNIKOVÉ PROCESY	Čas	Produkt	Sklad	Obchod	Zákazník
Nákup zboží	X	X	X		
Prodej zboží	X	X		X	X
Plán prodeje	X	X		X	

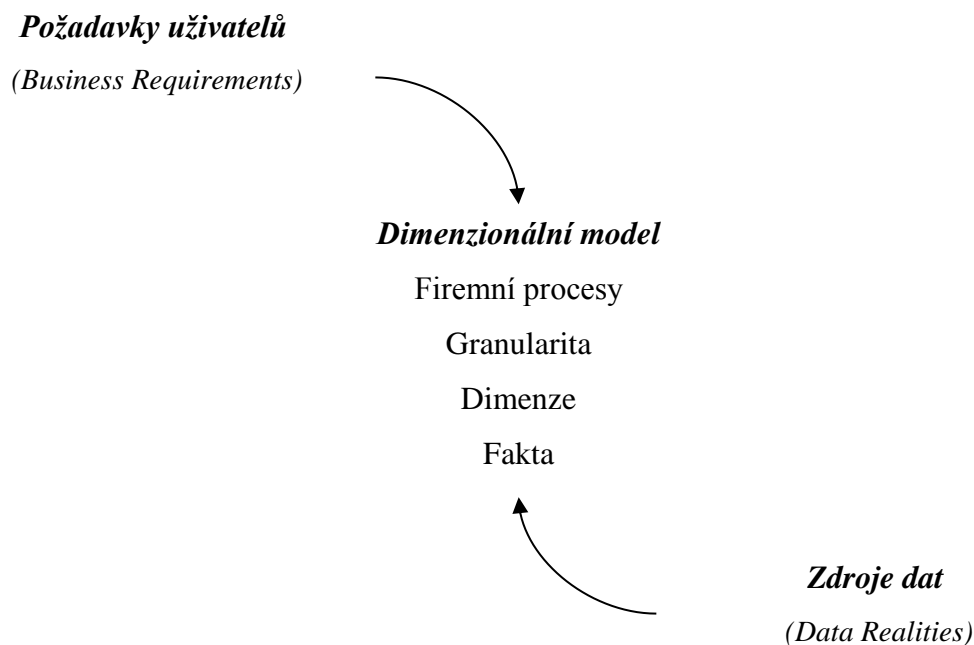
Zdroj: vlastní

Dalším krokem je stanovení granularity neboli úrovně detailu sledovaných procesů. Tím je řečeno na jaké podrobnosti budou dané fakty uchovávány a tím pádem i jak podrobnou analýzu bude možné provádět. V případech, kdy není možné granularitu faktu striktně určit, je doporučováno, navrhovat granularitu spíše vyšší než nižší, neboť dosažení vyšší agregace je vždy jednodušší než naopak.

Třetím krokem při návrhu je určení standardizovaných dimenzí. Ty musejí být pro splnění podmínky konformnosti vždy na nejvyšší úrovni granularity. Konformní dimenze využívané na různých úrovních granularity jsou poté pouze matematickou podmnožinou základní dimenze.

Posledním krokem je návrh faktů. Ty vyplývají z jednotlivých podnikových procesů a mohou být konformní. V závislosti na definici faktů mohou být definovány i konformní dimenze. Výskyt konformních dimenzí je v multidimenzionálním modelování naprosto běžný. Nejčastější konformní dimenzí je dimenze času.

Všechny tyto kroky vycházejí a také podléhají dvěma proudům informací. Jedním jsou požadavky uživatelů na určitou funkcionalitu a druhým jsou zdroje dat. Obrázek 3-6 tento proces znázorňuje. Jelikož některé uživatelské požadavky není možné kvůli neúplným nebo chybějícím datům splnit, je nutné tyto dva proudy sloučit a pomocí výše zmíněných kroků získat z jejich průniku výsledný datový model.



Obrázek 3-6: Proces návrhu dimenzionálního modelu

Zdroj: KIMBALL, Ralph a Margy ROSS. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd ed. New York: Wiley, c2002, s. 32, obr. 2.1. ISBN 0471200247.

4 Návrh řešení BI aplikace v prostředí společnosti BREX

V této kapitole se nachází hlavní cíl práce a to návrh řešení kompletní BI aplikace pro konkrétní potřeby společnosti B R E X, s.r.o. Tato kapitola se v první části zabývá popisem prostředí této společnosti. Dále je již kapitola členěna na několik kroků, kterým celý proces návrhu podléhá. Hlavními částmi řešení jsou: analýza požadavků na aplikaci, volba realizační platformy, návrh databázového modelu datového skladu, návrh procesu ETL, návrh datové kostky a návrh výsledných reportů.

4.1 Prostředí firmy

Společnost B R E X, s.r.o. (dále jen „BREX“) byla založena v roce 1991 skupinou odborníků z oblasti stavební výroby a stavební projekce. *„Rozložení těchto odborností předurčovalo zaměření firmy do širokého spektra stavebních oborů s možností provádět jak kompletní stavby, tak se specializovat na odborné technologie stavební výroby.“*²⁷

V roce 2013 se spoluvlastníkem společnosti stala společnost V - CON, s.r.o., součást skupiny VALBEK-EU, a.s.

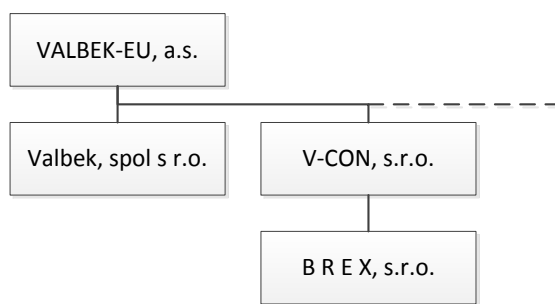
BREX se zabývá především výstavbou dopravních staveb, inženýrských sítí, průmyslových hal a pozemních či občanských staveb. Kromě toho nabízí i speciální práce, jako jsou například sanace betonových konstrukcí nebo hydroizolace. V současné době zaměstnává přibližně 150 zaměstnanců a její obrát se v roce 2013 pohyboval na hranici 350 milionů korun.²⁸

²⁷ BREX: Profil společnosti. B R E X, spol s. r.o. *BREX: stavební společnost* [online]. 2013 [cit. 2014-02-09]. Dostupné z: <http://www.brex.cz/cs/spolecnost/profil-spolecnosti/>

²⁸ BREX: Ekonomické informace. B R E X, spol s. r.o. *BREX: stavební společnost* [online]. 2013 [cit. 2014-02-09]. Dostupné z: <http://www.brex.cz/cs/spolecnost/ekonomicke-informace/>

4.1.1 Má pozice

Jak již bylo zmíněno, firma BREX patří do holdingu VALBEK-EU, jehož zjednodušená struktura je znázorněna na obrázku 4-1. Do stejné skupiny firem patří i společnost Valbek, spol. s r.o. (dále jen „Valbek“), ve které vykonávám roční řízenou praxi oboru manažerská informatika. Tato společnost se zabývá především projektováním inženýrských staveb a mostů. V rámci této praxe spadám pod středisko Aspe, které se zabývá vývojem stejnojmenného softwaru zaměřeného na investory, projektanty a dodavatele. Jedná se o ASW zprostředkovávající jednoduchou komunikaci mezi výše zmíněnými subjekty ohledně plánované, prováděné nebo dokončené stavební zakázky.



Obrázek 4-1: Struktura společnosti

Zdroj: vlastní

Jelikož projekční firma Valbek vidí ve spojení se stavební firmou BREX velký potenciál, jejich záměrem je vytvoření BI aplikace pro možnost analýzy dat z Aspe a libovolného ekonomického (účetního) systému současně.

Společnost BREX již nějakou dobu ASW Aspe používá pro evidenci zakázek a řízení stavebních prací. Pro evidenci účetnictví doposud využívá informační a řídicí systém IPOS²⁹, ale v současné době již dochází k postupnému přechodu na nový celopodnikový ERP systém – KARAT od společnosti KARAT Software a.s. Ke změně systému dochází především kvůli jednodušší konsolidaci účetních výkazů napříč všemi firmami patřící do holdingu VALBEK-EU, neboť všechny spadající pod tento holding již ERP KARAT

²⁹ IPOS-SOFT. *IPOS-SOFT: Komplexní informační a řídicí systémy pro investiční výstavbu* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.ipossoft.cz/>

používají. Vize projektu je tedy vytvoření BI aplikace sjednocující data z Aspe a KARATu pro jejich jednodušší analýzu.

4.2 Raná studie

V době nástupu na roční řízenou praxi, byl projekt pouze ve fázi velmi rané studie – spíše záměru projektu. Nebyl navržen datový model, který by odpovídal všem požadavkům. Aby bylo možné využít Kimballovu metodiku multidimenzionálního modelování, bylo rozhodnuto o provedení nové analýzy a tudíž projekt začít řešit od začátku.

4.3 Analýza požadavků

Při návrhu BI aplikace je stěžejní návrh datového modelu podnikového DWH. Při jeho návrhu je jedním ze základních úkolů sběr a analýza požadavků. Při analýze požadavků byly získané požadavky rozděleny do dvou oblastí, které jsou uvedeny dále.

4.3.1 Základní požadavky na aplikaci

Z několika diskuzí s vedoucími pracovníky vzešel seznam následujících požadavků na funkcionalitu BI aplikace.

- Rychlé výstupy s co nejkratší dobou odezvy založené na předpočítaných agregovaných datech.
- Možnost zadávat do BI aplikace různé verze plánovaných hodnot pro následující období a získávat výstupy porovnávající realitu s konkrétním plánem.
- Flexibilní řešení pro možnost budoucího rozšíření.
- Sledování účetních nákladů a výnosů z jiných úhlů pohledu než nabízí účetní software.
- Mapování nákladů a výnosů z ERP KARAT a položek kalkulačního vzorce (PKV) z Aspe do dynamické stromové struktury tvořící zjednodušenou účetní výsledovku pro možnost porovnání skutečného, provedeného a fakturovaného množství. Tuto strukturu firma nazývá „výsledkové skupiny“.

Tyto obecné požadavky udávají pouze charakter nebo směr koncepce celé aplikace. Pro získání konkrétnějších odpovědí, dle kterých bylo možné sestavit datový model, bylo potřeba se ptát dále a mnohem detailněji.

4.3.2 Požadavky na dotazy

Nejzákladnější otázkou při sběru konkrétních požadavků na výstupy z aplikace je: „*Na jaké otázky chcete znát odpovědi?*“³⁰ Zjištění všech požadavků a nároků od pracovníků, kteří budou DWH využívat, je prvním klíčovým bodem při návrhu celé BI aplikace. V mnohých případech jsou ale některé požadavky nesplnitelné především z důvodu neúplných nebo chybějících dat. Proto je třeba hledat kompromis mezi získanými požadavky a dostupnými daty. Důležité také je, zamyslet se nad ekonomickou stránkou řešení a přínosem pro firmu. Zejména pokud bude konkrétní požadavek důležitý pro podporu rozhodování či nikoliv. Stejný postup jsem využil při sběru požadavků i já. Z dalších diskuzí s vedoucími pracovníky vzešel následující konkrétnější seznam požadavků.

- Sledovat změny během výstavby (ZBV) rozdělené na méněpráce a vícepráce na konkrétních položkách rozpočtu (dále jen „položky“) dle času a kategorií ve kterých jsou zařazeny. Pokud není kategorie položky vyplněna, jedná se o kategorii „Neznámá“.
- Sledovat skutečné čerpání položek v čase za jednotlivé PKV.
- Sledovat skutečně fakturované částky za stavební objekty (dále jen „objekty“) v čase a PKV.
- Sledovat skutečně zaúčtované částky z účetního systému v čase a dělit je dle nákladových a výnosových skupin.
- Sledovat rozdíl fakturovaných a zaúčtovaných částek v závislosti na výsledkových skupinách.

³⁰ WITHEE, Ken. *Microsoft Business Intelligence For Dummies*. Hoboken, NJ: Wiley Pub., 2010, s. 63. ISBN 978-0-470-52693-4.

- Zadávat plánované hodnoty na jednotlivé výsledkové skupiny pro různé útvary podniku za měsíční interval.
- Sledovat skutečné čerpání stavebních objektů oproti plánovanému.
- Definovat několik různých scénářů celopodnikových plánů pro jednotlivé nákladové a výnosové skupiny, které bude možné porovnávat s realitou.

Dalším požadavkem u většiny procesů je sledování v čase na úrovni měsíční podrobnosti a v závislosti na firemním útvaru.

4.4 Analýza zdrojů

Jak již bylo v předešlé kapitole zmíněno, firma chce pomocí BI aplikace analyzovat data z výrobního ASW Aspe a ekonomické (účetní) části ERP KARAT. Toto jsou tedy jediné dva transakční systémy, které budou sloužit jako zdroje dat.

4.4.1 Aspe

„Aspe je software, který je určen pro přípravu a realizaci staveb. Funguje jako jestřábí oko, které vám poskytne souhrnný přehled o všech plánovaných a realizovaných stavbách.“³¹

Jedná se o software pro řízení stavebně investičních projektů, pokrývající všechny činnosti od přípravy přes výběrové řízení až po detailní sledování realizace. Jedinečný softwarový nástroj pro investory, projektanty a dodavatele staveb, který je nutností pro efektivní řízení stavebních projektů.

Jelikož je Aspe na správu zakázek zaměřen a řešení je navrhováno pro stavební firmu, téměř všechna data o zakázkách budou pocházet z Aspe. Pro ukládání dat využívá vlastní relační databáze na platformě MS SQL Server. Konkrétní data, která budou pocházet z této databáze, budou zmíněna v následujících kapitolách.

³¹ VALBEK. Co je Aspe. *Aspe: Software pro řízení staveb* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.aspe.cz/cs/produkty/aspe/co-je-aspe>

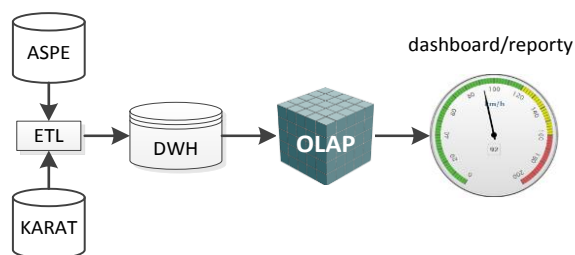
4.4.2 KARAT

Informační systém KARAT od společnosti KARAT Software a.s. je všestranný podnikový ERP systém. Jeho široká škála nabízených modulů s možností velmi detailního nastavení dokáže komplexně pokrýt velké množství firemních procesů. Jelikož firma BREX tento program využívá pro vedení účetnictví, data účetních výkazů budou pocházet z tohoto systému. Stejně jako Aspe i KARAT používá pro ukládání dat MS SQL Server.³²

Jedním z nabízených produktů firmy KARAT je i BI nadstavba nad jejich ERP systémem. Pomocí něj je však možné analyzovat pouze ekonomická data pocházející z této aplikace. Jelikož BREX potřebuje především analýzu dat z Aspe, která je doplněna o účetní údaje, pro firmu není výhodné tento modul využít.

4.5 Volba realizační platformy

Po analýze požadavků a zdrojů bylo možné určit koncept aplikace a stanovit, jak by měl výsledek vypadat. Na základě nasbíraných požadavků bylo zvoleno standardní řešení BI skládající se z DWH, ETL, OLAP databáze a reportů. Schématické znázornění navrženého systému se nachází na následujícím obrázku.



Obrázek 4-2: Koncept BI aplikace společnosti BREX

Zdroj: vlastní

BI aplikace je možné realizovat na různých softwarových platformách. Na ně bývají od zadavatele kladeny různé nároky. Mezi nejběžnější patří nízké pořizovací náklady, malé náklady spojené s údržbou a co nejmenší zásah do stávající podnikové IT infrastruktury.

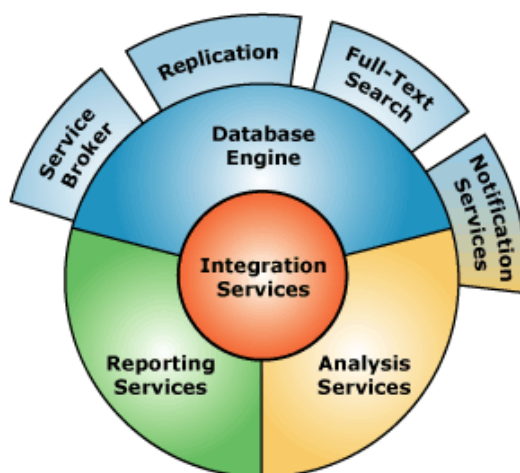
³² KARAT SOFTWARE A. S. *Karat: informační systém* [online]. 2014 [cit. 2014-05-03]. Dostupné z: <http://www.karatsoftware.cz/>

Pokud je v podniku zvažováno nasazení BI technologií, ideální možností je využití již stávajícího vybavení. To je ovšem možné pouze za předpokladu, že dostupné prostředky podporují všechny potřebné technologie.

Takovýto „ideální“ scénář se nachází i ve společnosti BREX. Ta již disponuje databázovým serverem s licencí na MS SQL Server 2008 R2 Enterprise, která splňuje všechny požadavky na realizační platformu plánované BI aplikace. Kromě výrazně nižších pořizovacích nákladů, než při pořízení nové licence, je tato platforma pro realizaci BI řešení hojně využívána a nabízí, kromě jiného, kvalitní produktovou dokumentaci a širokou podporu. Díky všem těmto kladům se firma rozhodla pro využití stávající platformy.

4.5.1 Microsoft SQL Server 2008 R2

Tento softwarový produkt od společnosti Microsoft je zaměřen na správu databází, skladování a dolování dat³³. Celý program je členěn na čtyři hlavní části – viz obrázek 4-3, které dohromady tvoří mocný nástroj na tvorbu BI aplikací.



Obrázek 4-3: Moduly MS SQL Serveru 2008 R2

Zdroj: <http://i.technet.microsoft.com/dynimg/IC102493.gif>

Stěžejním modulem je databázový server pro správu relačních databází (Database Engine). Pro analýzu dat je zde druhý modul – analytický server – SQL Server Analysis Services

³³ MICROSOFT. *Microsoft SQL Server* [online]. 2012 [cit. 2014-04-18]. Dostupné z: <http://www.microsoft.com/sqlserver/cs/cz/default.aspx>

(SSAS). Ten umožňuje provádět hromadnou analýzu dat pomocí OLAP databází s datovými kostkami. Reportingový server SQL Server Reporting Services (SSRS) umožňuje vytvářet přehledné grafické reporty z dat umístěných v OLAP databázích, relačních databázích nebo jiných zdrojů. Přenosy dat mezi těmito moduly a dalšími externími zdroji dat zajišťuje SQL Server Integration Services (SSIS). Tato služba slouží především jako nástroj pro vytváření ETL procesů.

4.6 Návrh datového modelu databáze datového skladu

Po důkladné analýze a volbě realizační platformy bylo možné pokročit k dalšímu kroku, jímž je návrh datového modelu. Pro vytvoření datového modelu byla využita Kimballova metoda multidimenzionálního modelování, jež pro analýzu faktů a dimenzí používá sběrníkovou architekturu. Ze získaných požadavků bylo možné takovou BUS matici sestavit – viz tabulka 4-1 (s. 47).

Z této matice je možné vyčíst konformní dimenze a fakty, avšak jejich popis není dostatečně detailní, aby se z nich dal sestavit fyzický datový model. Je potřeba každou dimenzi detailněji rozpracovat a specifikovat jaké má obsahovat atributy, hierarchie a vztahy mezi atributy. Detailně rozpracované dimenze se nacházejí v podkapitole 4.6.1.

Standardní BUS matice znázorňuje pouze procesy a dimenze. Pro lepší znázornění byla matice vylepšena a jsou v ní znázorněny také konkrétní fakty spojené s daným procesem, a také na jaké úrovni granularity jsou dané fakty s dimenzí spojeny.

Tabulka 4-1: BUS matice podnikového datového skladu společnosti BREX

PROCESY		FAKTY		DIMENZE									
				Čas	Partneři	Útvary	Stavby		Kategorie ZBV	Dodatky	Zjišťovací protokoly	Scénáře	Potřeby
Obraty staveb v účetnictví		X	X	X	Objekt	Položka							X
	strana MD	X	X	X	X								
	strana DAL	X	X	X	X								X
Plán stavebních zakázek	částka	X	X	X	X						X		X
	cena původní			X	X	X							
Rozpočet staveb	cena nákladová			X	X	X							
	cena aktuální			X	X	X							
Skutečná fakturace staveb v Aspe	částka	X	X	X	X	X				X		X	X
	cena	X	X	X	X	X		X	X				
Čerpání staveb	množství	X	X	X	X	X		X	X				
	méně práce	X		X	X	X		X					
Změny během výstavby	více práce	X		X	X	X		X	X				
	cena			X	X	X		X	X				
Rozpočet ZBV	množství			X	X	X		X	X				
	částka	X		X							X		X
Plán firmv													

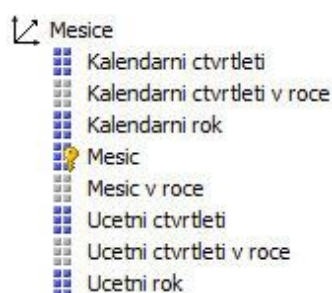
Zdroj: vlastní

4.6.1 Dimenze

V návrhu tohoto DWH se nachází celkem 10 dimenzí, z čehož jsou, až na jednu výjimku, všechny konformní. V následujících podkapitolách jsou detailně popsány atributy, jejich vztahy a hierarchie všech dimenzí obsažených v návrhu. Nejdůležitější z nich je časová dimenze.

4.6.1.1 Čas

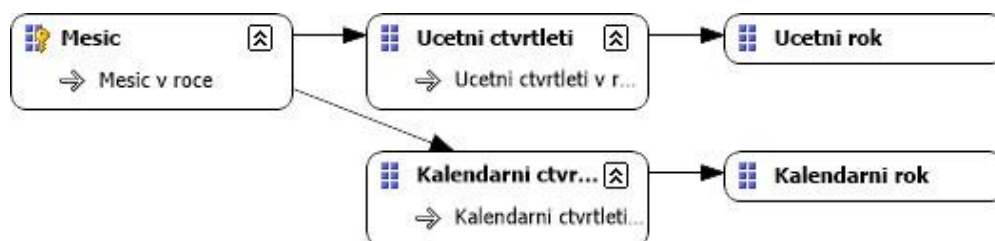
Navržená časová dimenze je na měsíční podrobnosti a obsahuje informace o kalendářním a účetním roce, kalendářním a účetním čtvrtletí a měsíci. Všechny dostupné atributy jsou na následujícím obrázku. Modře zvýrazněné jsou úrovně atributy a šedivě popisné.



Obrázek 4-4: Atributy časové dimenze

Zdroj: vlastní

Vztahy mezi jednotlivými atributy jsou definovány následujícím způsobem.



Obrázek 4-5: Vztahy mezi atributy v časové dimenzi

Zdroj: vlastní

V rámci časové dimenze jsou dostupné dvě hierarchie a to dle kalendářního a účetního roku. Jejich náhled je na následujícím obrázku.



Obrázek 4-6: Hierarchie v časové dimenzi

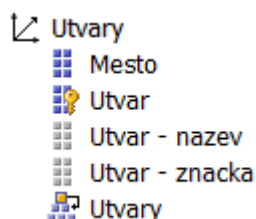
Zdroj: vlastní

Dimenze obsahuje nejen numerické označení roků, kvartálů a měsíců, ale také jejich jednoznačný popis. Popis roků je ve formátu „K2014“ pro kalendářní rok a „Ú2014“ pro

účetní. Čtvrtletí pak mají obdobný formát. „Q1 K2014“ a „Q1 Ú2014“. Měsíce jsou společné pro obě hierarchie a jejich popis je ve tvaru „leden 2014“. Data této dimenze jsou v DWH předem vygenerována, neboť všechny její atributy jsou předem dobře známe. Pokud bude dimenze vygenerována na dostatečně dlouho dobu dopředu (např. 20 let), není nutné se o ni dále jakkoliv starat.

4.6.1.2 Útvary

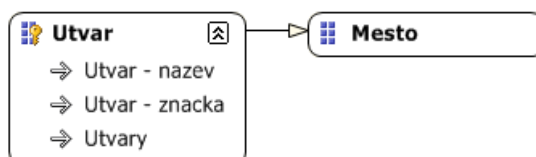
Dimenze podnikových útvarů je typickou ukázkou parent-child dimenze, neboť je potřeba určit hierarchickou strukturu firemních oddělení. Atributy této dimenze jsou vidět na následujícím obrázku.



Obrázek 4-7: Atributy v dimenzi útvarů

Zdroj: vlastní

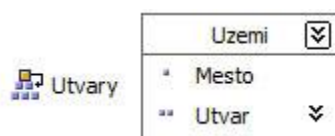
Vztahy mezi jednotlivými atributy jsou definovány následujícím způsobem.



Obrázek 4-8: Vztahy mezi atributy v dimenzi útvarů

Zdroj: vlastní

Tato dimenze obsahuje dvě hierarchie. Strukturu firemních útvarů a rozdělení dle území.



Obrázek 4-9: Hierarchie v dimenzi útvarů

Zdroj: vlastní

Jelikož je struktura firemních útvarů definována v Aspe i KARATu, není jednoznačně určeno, odkud budou data pro tuto dimenzi pocházet. Touto problematikou se dále zabývá kapitola věnující se návrhu ETL. V rámci této dimenze je definován i výchozí prvek s číslem (ID) 0 a názvem „Neznámý“. Ten bude využit při plnění tabulek faktů. Pokud nebude odkaz na daný útvar nalezen, bude automaticky nahrazen za tento výchozí.

4.6.1.3 Stavby

Další stěžejní dimenzí je dimenze staveb, která je sdílena téměř mezi všemi fakty. Ta je bezesporu největší a nejobjemnější dimenzí, která se v DWH vyskytuje. Obsahuje informace o jednotlivých stavbách neboli stavebních zakázkách. Všechny atributy jsou uvedeny na obrázku níže.

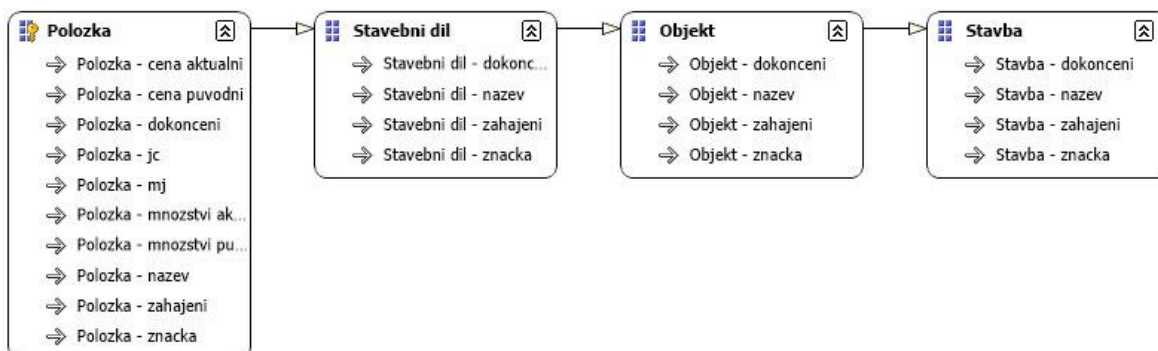
Stavby	Objekt	Stavebni dil	Polozka	Polozka - mj
Stavba	Objekt - dokonceni	Stavebni dil - dokonceni	Polozka - cena aktualni	Polozka - mnozstvi aktualni
Stavba - dodavatel	Objekt - nazev	Stavebni dil - nazev	Polozka - cena puvodni	Polozka - mnozstvi puvodni
Stavba - dokonceni	Objekt - zahajeni	Stavebni dil - zahajeni	Polozka - dokonceni	Polozka - nazev
Stavba - investor	Objekt - znacka	Stavebni dil - znacka	Polozka - jc	Polozka - zahajeni
Stavba - nazev				Polozka - znacka
Stavba - zahajeni				
Stavba - znacka				

Obrázek 4-10: Atributy v dimenzi staveb

Zdroj: vlastní

Každý z úrovněvých atributů k sobě váže několik popisných atributů. Všechny obsahují značku, název a datum zahájení a dokončení. Stavby mají navíc atribut určující její stav, který může nabývat hodnot „prováděná“ nebo „dokončená“, dále také jméno investora a dodavatele konkrétní stavby. U položek je navíc uvedena jednotková cena, původní cena, aktuální cena a původní a aktuální množství.

Vztahy mezi uvedenými atributy jsou znázorněny na následujícím obrázku.



Obrázek 4-11: Vztahy mezi atributy v dimenzi staveb

Zdroj: vlastní

Tato dimenze obsahuje pouze jednu hierarchii, kterou je samotná struktura staveb. Ta je členěna na čtyři úrovně. Pod stavby spadají stavební objekty, pod nimi jsou stavební díly a nejnižší úrovní jsou položky – viz obrázek 4-12.

Stavby	
* Stavba	▼
** Objekt	▼
*** Stavebni dil	▼
**** Polozka	▼

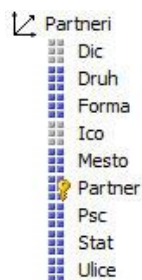
Obrázek 4-12: Hierarchie v dimenzi staveb

Zdroj: vlastní

Všechna data této dimenze pocházejí z Aspe.

4.6.1.4 Partneri

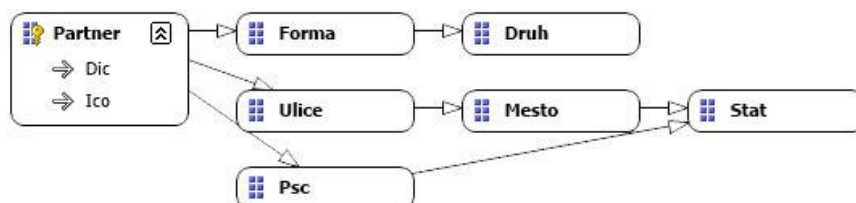
Tato dimenze obsahuje seznam obchodních partnerů. Každý záznam by měl obsahovat základní informace o firmě, jako jsou název, adresa, IČO a DIČ. Dále dimenze obsahuje informace o tom, o jaký druh společnosti se jedná – zda jde o fyzickou či právnickou osobu a jakou formu společnosti se jedná (s. r. o., a. s. atd.). Všechny definované atributy jsou znázorněny následujícím obrázkem.



Obrázek 4-13: Atributy v dimenzi partnerů

Zdroj: vlastní

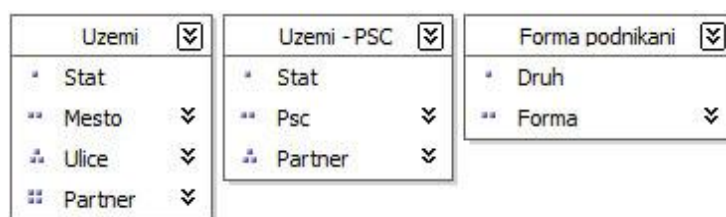
Vztahy mezi atributy v této dimenzi jsou poměrně složité a jsou znázorněny na následujícím obrázku.



Obrázek 4-14: Vztahy mezi atributy v dimenzi partnerů

Zdroj: vlastní

Z adresních atributů lze sestavit hierarchii dle geografického členění. Ta může mít dvě struktury, které se liší použitím PSČ. Dimenze ještě obsahuje hierarchii dle formy podnikání. Obrázek 4-15 tyto hierarchie dokumentuje.



Obrázek 4-15: Hierarchie v dimenzi partnerů

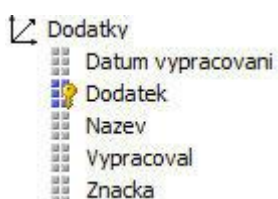
Zdroj: vlastní

Kvůli častému vzniku duplicit v transakčních systémech je třeba dávat velký pozor na to, jaká data se do dimenze dostávají. Duplicity vedou k nepřesným výstupům a chybné analýze. U této dimenze také není možné přesně určit, odkud budou data čerpána, neboť seznamy obchodních partnerů jsou opět vedeny v obou zdrojových systémech. Problematika zdroje dat pro tuto dimenzi je opět uvedena dále. Tato dimenze také obsahuje

výchozí prvek s číslem 0 a názvem „Neznámý“. Jeho použití je stejné jako u dimenze firemních útvarů.

4.6.1.5 Dodatky

Dodatek stavby určuje informace ohledně změny oproti plánovanému množství nebo ceně stavební práce. Stavba může obsahovat několik takovýchto dodatků a tato dimenze je uchovává. Každý záznam obsahuje informace o názvu, dni vypracování a osobě, která daný dodatek zpracovala. Všechny atributy jsou uvedeny na následujícím obrázku.



Obrázek 4-16: Atributy v dimenzi dodatků

Zdroj: vlastní

Vztahy mezi jednotlivými atributy jsou definovány následujícím způsobem.



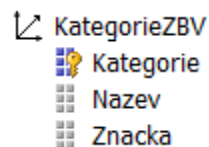
Obrázek 4-17: Vztahy mezi atributy v dimenzi dodatků

Zdroj: vlastní

V rámci dimenze dodatků nejsou definovány žádné hierarchie. Data této dimenze opět pocházejí z Aspe.

4.6.1.6 Kategorie ZBV

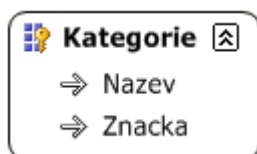
Dimenze kategorií ZBV je velmi malou dimenzí. Její předpokládaná velikost je odhadnuta na maximálně několik desítek záznamů. Obsahuje pouze informace o čísle (značce) a názvu kategorie, která definuje typ změny v projektu oproti plánu. Atributy dimenze znázorňuje obrázek 4-18.



Obrázek 4-18: Atributy dimenze kategorií ZBV

Zdroj: vlastní

Vztahy mezi jednotlivými atributy jsou definovány následujícím způsobem.



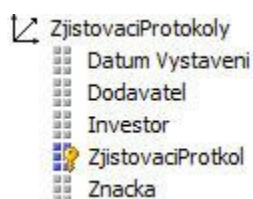
Obrázek 4-19: Vztahy mezi atributy v dimenzi kategorií ZBV

Zdroj: vlastní

Definování jakýchkoliv hierarchií zde není možné a postrádalo by smysl. Kategorie ZBV jsou záležitostí čistě stavebního charakteru, proto také data této dimenze pocházejí z Aspe. Stejně jako předešlá dimenze i tato obsahuje výchozí prvek.

4.6.1.7 Zjišťovací protokoly

Zjišťovací protokoly (ZP) se používají při komunikaci mezi investorem a dodavatelem ohledně skutečně provedeného a fakturovaného množství (peněz). Díky této dimenzi je možné určit, kolik bylo daným ZP čerpáno. Atributy dimenze jsou znázorněny na následujícím obrázku.



Obrázek 4-20: Atributy v dimenzi zjišťovacích protokolů

Zdroj: vlastní

Jelikož dimenze obsahuje kromě klíče pouze popisné atributy, jejich vztah je velmi jednoduchý – viz obrázek 4-21.



Obrázek 4-21: Vztahy mezi atributy v dimenzi zjišťovacích protokolů

Zdroj: vlastní

Tato dimenze neobsahuje žádné hierarchie. Opět se jedná o dimenzi pocházející z Aspe.

4.6.1.8 Scénáře

Pro možnost definice různých verzí plánovaných hodnot a jejich případné porovnání je zde dimenze možných scénářů. Jedná se opět o drobnou dimenzi obsahující pouze název plánu, verzi a datum schválení. Ani jeden atribut zde není agregovatelný, neboť sčítat hodnoty za všechny scénáře postrádá smysl. Zmíněné atributy a jejich vztahy jsou vyobrazeny na následujících dvou obrázcích.



Obrázek 4-22: Atributy v dimenzi scénářů

Zdroj: vlastní



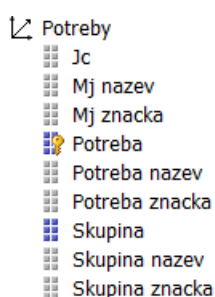
Obrázek 4-23: Vztahy mezi atributy v dimenzi scénářů

Zdroj: vlastní

Tato dimenze neobsahuje žádné hierarchie a její data nepocházejí ani z jedné ze zdrojových aplikací. Musejí být do DWH zadána ručně nebo přes nějakou vlastně vytvořenou aplikaci pro správu podnikových plánů.

4.6.1.9 Potřeby

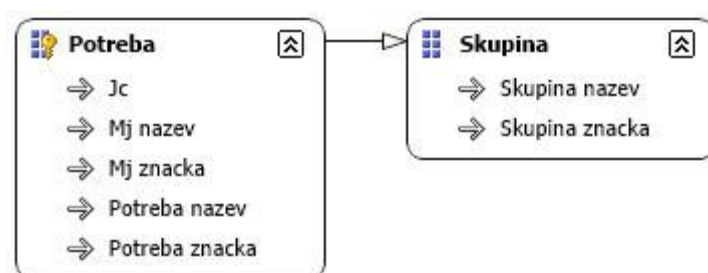
Každá položka stavby se může skládat z různých potřeb. Těmi jsou například materiály, mzdy za odvedenou práci, poplatky za stroje atd. Aby bylo možné zjistit, jaké celkové množství všech potřeb bylo na stavbu vynaloženo, je zde tato dimenze. Potřeba má vždy jednotkovou cenu, měrnou jednotku a patří do nějaké skupiny potřeb. Základními skupinami jsou hmoty, mzdy a stroje, ale může jich být libovolné množství. Všechny atributy dimenze potřeb jsou na následujícím obrázku.



Obrázek 4-24: Atributy v dimenzi potřeb

Zdroj: vlastní

Vztahy mezi atributy skupin a potřeb jsou vidět na obrázku níže.



Obrázek 4-25: Vztahy mezi atributy v dimenzi potřeb

Zdroj: vlastní

V rámci této dimenze se vyskytuje jedna hierarchie nazvaná „Potřeby“. První úrovní jsou skupiny potřeb a druhou pak samotné potřeby – viz obrázek 4-26.

Potřeby	☑
▪ Skupina	▼
▪▪ Potřeba	▼

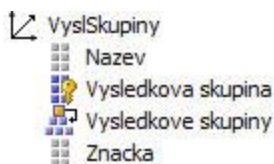
Obrázek 4-26: Hierarchie v dimenzi potřeb

Zdroj: vlastní

Dimenze potřeb je plněna z databáze Aspe, ale její plnění a aktualizace jsou poměrně problematické. Potřeby totiž nemají v databázi Aspe žádný identifikátor, neboť jsou uživatelsky ručně zadávány na konkrétních položkách rozpočtu samostatně. Z toho důvodu je například jedna potřeba pro čtyři různé položky v databázi vedena 4x. Identifikátor je tedy při přenosu dat generován automaticky, ale při každém dotazu na tato data může být jiný, což značně komplikuje aktualizaci záznamů této dimenze. Více ohledně plnění tabulek dimenzí je uvedeno v podkapitole 4.7. Výchozí prvek je definován i v této dimenzi.

4.6.1.10 Výsledkové skupiny

Jak již bylo řečeno v úvodní analýze, výsledkové skupiny tvoří svou stromovou strukturou zjednodušenou účetní výsledovku. Proto i struktura této dimenze musí odpovídat parent-child dimenzi. Tato dimenze obsahuje pouze značku a název a odkaz na předchůdce – viz následující obrázky.



Obrázek 4-27: Atributy v dimenzi výsledkových skupin

Zdroj: vlastní



Obrázek 4-28: Vztahy mezi atributy v dimenzi výsledkových skupin

Zdroj: vlastní

Tato dimenze obsahuje pouze jedinou hierarchii, výsledková skupina. V rámci každého záznamu dimenze je také definován operátor (+ nebo -), který je na skupinu aplikován při agregaci v rámci stejné úrovně. Tímto je možné docílit toho, že některé „větve“ stromu se při konečném součtu odečtou. Například odečtení nákladů od výnosů pro získání zisku.

Stejně jako u dimenze scénářů ani data této dimenze nepocházejí ani z jednoho zdrojového systému. Data je potřeba do dimenze nahrát přímo prostřednictvím DWH nebo jiné aplikace.

4.6.2 Fakty

Návrh analyzuje celkem 8 firemních procesů, ze kterých následně vyplývá 14 fyzických faktů – viz tabulka 4-1 (s. 47). Tyto procesy a fakty jsou detailněji popsány v následující tabulce.

Tabulka 4-2: Detailnější popis definovaných faktů

Název procesu	Fakt	Popis	Typ ³⁴	Agr. funkce
Obraty staveb v účetnictví	strana MD	Zaúčtované hodnoty stavebních zakázek v účetnictví – strana má dáti (MD).	A	SUM
	strana D	Zaúčtované hodnoty stavebních zakázek v účetnictví – strana dal (D).	A	SUM

³⁴ A – aditivní, SA – semi-aditivní, NA – neaditivní

Název procesu	Fakt	Popis	Typ ³⁴	Agr. funkce
Plán stavebních zakázek	částka	Plánované náklady a výnosy na stavbu, (možné a získané zakázky) Fakt nelze agregovat za dimenzi scénářů.	SA	SUM
Rozpočet staveb	cena původní	Původní cena rozpočtu stavby dle smlouvy o díle (SOD).	A	SUM
	cena nákladová	Aktuální nákladová cena stavby.	A	SUM
	cena aktuální	Aktuální cena rozpočtu stavby navýšená o plánované ZBV.	A	SUM
Skutečná fakturace staveb v Aspe	částka	Skutečná cena stavby dle výrobních faktur z Aspe.	A	SUM
Čerpání staveb	cena	Čerpaná částka staveb z Aspe	A	SUM
	množství	Čerpané množství staveb z Aspe.	SA	SUM
		Fakt je označen jako semi-aditivní z důvodu postrádání smyslu při agregaci za dimenzi staveb.		
Změny během výstavby	méněpráce	Cena provedených méněprací v rámci ZBV	A	SUM
	vícepráce	Cena provedených víceprací v rámci ZBV	A	SUM
Rozpočet ZBV	cena	Ceny plánovaných změn během výstavby.	A	SUM
	množství	Množství plánovaných změn během výstavby. Agregace za dimenzi staveb postrádá smysl.	SA	SUM
Plán firmy	částka	Plánované náklady/výnosy firmy v budoucích obdobích. Fakt nelze agregovat za dimenzi scénářů.	SA	SUM

Zdroj: vlastní

Všechny fakty spadající pod jeden proces jsou navzájem konformní, neboť využívají stejnou množinu dimenzí. Další fakty, které mohou být s použitím těchto fyzických faktů vypočteny, jsou dále rozebrány v kapitole věnující se návrhu datové kostky.

4.6.3 Fyzický model databáze

Asi nejdůležitější částí celého projektu je vytvoření fyzického datového modelu databáze datového skladu. Ten je zpracován rovněž v rámci MS SQL Serveru. Model zcela odpovídá návrhu všech dimenzí a faktů popsaných výše. Datové typy atributů byly převzaty z databází, odkud dané atributy pocházejí. Všechny atributy sloužící jako PK jsou datového typu INT kvůli dostatečné mohutnosti a rychlosti indexace. Všechny faktové atributy nesoucí údaj o peněžích jsou datového typu MONEY a atributy nesoucí informaci o množství pak typu DECIMAL(18,6). Všechny tabulky faktů mají PK složený ze všech FK odkazujících na tabulky dimenzí. To zajistí opravdu velkou mohutnost PK a téměř nemožnost vyčerpání počtu kombinací.

Na následujícím obrázku je návrh databázové tabulky obsahující fakty o čerpání staveb. Název tabulky je „*fact_stavby_cerpani*“, uchovává 4 fyzické fakty a její PK tvoří 7 FK.

fact stavby cerpani			
	Column Name	Data Type	Allow Nulls
🔑	id_dim_mesice	int	<input type="checkbox"/>
🔑	id_dim_partneri	int	<input type="checkbox"/>
🔑	id_dim_utvary	int	<input type="checkbox"/>
🔑	id_dim_polozky	int	<input type="checkbox"/>
🔑	id_dim_kategorie_zbv	int	<input type="checkbox"/>
🔑	id_dim_dodatky	int	<input type="checkbox"/>
🔑	id_dim_zjistovaci_protokoly	int	<input type="checkbox"/>
	castka	money	<input type="checkbox"/>
	mnozstvi	numeric(19, 6)	<input type="checkbox"/>
	castka_provedeno	money	<input type="checkbox"/>
	mnozstvi_provedeno	numeric(19, 6)	<input type="checkbox"/>
			<input type="checkbox"/>

Obrázek 4-29: Návrh tabulky *fact_stavby_cerpani*

Zdroj: vlastní

Obrázek 4-30 zachycuje návrh tabulek dimenzí staveb. Z popisu dimenzí vyplývá, že dimenze staveb je logicky členěna do čtyř úrovní; stavby, objekty, stavební díly a položky. Jelikož každý z úrovnových atributů obsahuje také několik popisných, bylo rozumné tuto dimenzi normalizovat, což zmenší její celkovou velikost a zpřehlední ji. Dalším důvodem k normalizaci je také fakt, že některé fakty je potřeba napojit do této dimenze na vyšší

úrovni granularity, než na nejnižších položkách. Pro zachycení této vazby již na úrovni databáze je potřeba mít dimenzi rozdělenou do dalších logických celků.



Obrázek 4-30: Návrh tabulek dimenzí staveb

Zdroj: vlastní

Výše popsané tabulky tvoří pouze zlomek návrhu celé databáze. Celý návrh fyzického modelu databáze se nachází v přílohách A a B.

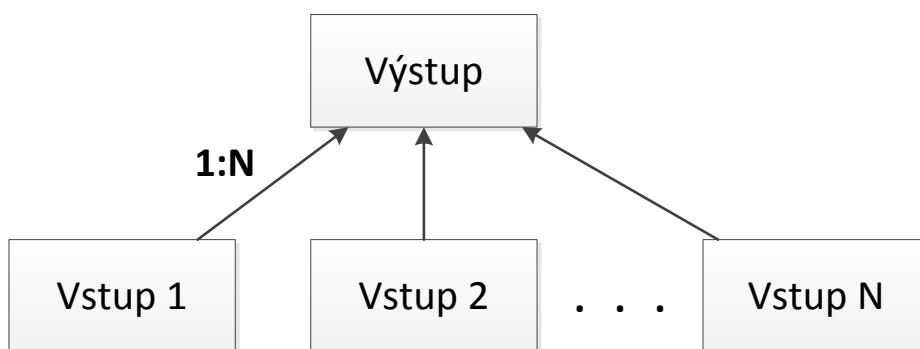
4.7 Návrh ETL

Ve fázi, kdy je hotový logický i fyzický model DWH, je třeba se začít zabývat tím, jakým způsobem se budou data do DWH dostávat, neboli navrhnout ETL procesy. Analýza zdrojů již byla provedena a zdroji jsou v tomto případě dvě transakční aplikace. Detailní informace o stavební části pocházejí z Aspe a ekonomické údaje z KARATu. Je tedy potřeba načíst data z těchto systémů, transformovat je a uložit do DWH. Nejprve se kapitola bude zabývat plněním dimenzí a poté tabulek faktů.

4.7.1 Plnění dimenzí

Z návrhu vyplývá, že některé konformní dimenze jsou sdíleny fakty, které pocházejí z různých systémů. Takové dimenze nelze plnit jednoduchým způsobem. Stejné záznamy bývají v různých systémech vedeny pod různými identifikátory a jejich atributy mívají i rozdílné datové typy, což aktualizaci takovýchto dimenzí značně komplikuje. Tím vzniká hned několik problémů. První otázkou je, jak takovou dimenzi naplnit, jestliže má být sjednocením unikátních výskytů stejných entit z různých systémů. A dále jak takovou dimenzi aktualizovat a jak zpracovávat duplicity. Tyto problémy se zpravidla týkají všech dimenzí, které jsou sdíleny mezi více systémy. Těmi jsou konkrétně v tomto případě dimenze obchodních partnerů, útvarů, výsledkových skupin, na které jsou mapovány PKV z Aspe a nákladové/výnosové účty z KARATu a dimenze staveb z Aspe, na které jsou zas mapovány zakázky z KARATu.

Pokud je třeba zajistit nějakou synchronizaci dat napříč systémy, je na místě využití MDM – viz kapitola 2.2.3. Klasickou metodu MDM lze ale prakticky využít pouze tehdy, pokud zúčastněné systémy podporují nějakou formu externí synchronizace dat. To bohužel ani jeden z firmou využívaných systému nepodporuje, tudíž bylo nutné navrhnout jinou metodu, jak sjednocení a standardizaci údajů zajistit. Byla tedy navržena metoda tzv. „mapování“. Ta spočívá v definici jednoho (centrálního) výstupního seznamu a libovolného počtu vstupních, které u sebe nesou informaci o vazbě na výstupní seznam. Funkce je znázorněna na následujícím schématu.



Obrázek 4-31: Schéma funkce mapování záznamů

Zdroj: vlastní

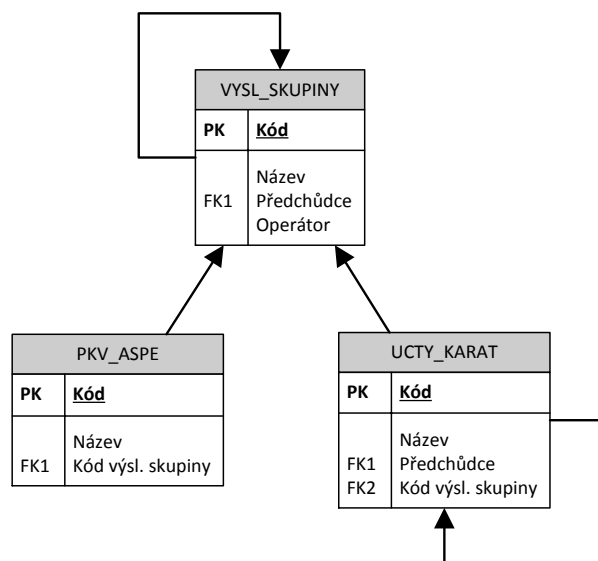
Tato metoda umožňuje vytvořit vazby mezi libovolným počtem zdrojových systémů, odstranit duplicity ze zdrojových dat díky vazbě 1:N mezi výstupním a vstupním seznamem a mít tak data, která se do dané dimenze z výstupního seznamu dostanou plně pod kontrolou. Nevýhodou tohoto řešení je naopak nemožnost automatické aktualizace výstupního seznamu na základě změny ve vstupních a nemožnost zpětného zápisu dat do zdrojových systémů.

4.7.2 SQL Server Master Data Services

Od verze 2008 R2 nabízí MS SQL Server nástroj pro realizaci MDM zvaný Master Data Services (MDS), který je však součástí pouze dvou nejvyšších nabízených verzí, jimiž jsou edice Datacenter a Enterprise. Tyto verze jsou ovšem využívány spíše velkými korporacemi. Firmy podobného rozsahu jako má společnost VALBEK-EU mnohem častěji využívají spíše edice Standard. Jelikož SQL Server 2008 R2 ve verzi Enterprise nabízí tento nástroj pro realizaci MDM, který je dostatečně flexibilní pro zajištění funkce výše zmíněné metody, a firma edicí Enterprise disponuje, rozhodli jsme se tento nástroj využít.

Tento systém pro realizaci MDM má několik základních funkcí, díky kterým je možné nadefinovat libovolné datové modely s rozličnými typy atributů. Při vytvoření nového projektu je nejprve nutné vytvořit nový datový model. Pod ním je možné vytvářet jednotlivé datové entity. Ty reprezentují konkrétní číselníky dat, podobně jako databázové tabulky. U každé entity je možné vytvořit téměř libovolný počet atributů. Ty mohou být definovány pomocí jednoho z pěti nabízených datových typů. Základními jsou textový atribut se zadanou maximální délkou a numerický atribut s definovaným počtem desetinných míst. Dále může být atribut typu soubor nebo odkaz na webovou stránku začínající znaky „http:/“. Posledním možným typem je tzv. doménový atribut. Jedná se o atribut, který odkazuje na jinou entitu ve stejném modelu. Princip je stejný jako u vazeb FK na PK v databázi.

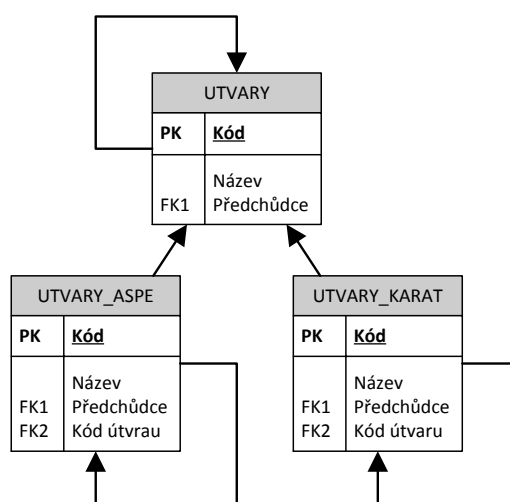
Díky možnosti doménových atributů je možné jednoduše vytvořit datový model pro navrženou metodu mapování. Pro potřeby tohoto DWH byl prostřednictvím MDS navržen následující datový model pro správu sdílených dimenzí.



Obrázek 4-32: Návrh datového modelu MDS pro mapování výsledkových skupin

Zdroj: vlastní

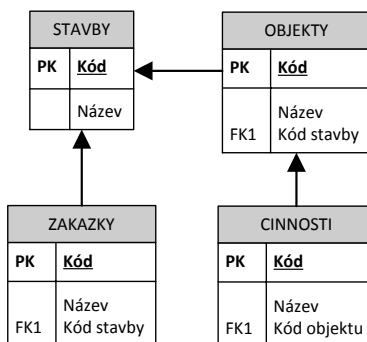
Obrázek 4-32 znázorňuje model pro párování výsledkových skupin, PKV z Aspe a účtů z KARATu. Všechny entity obsahují unikátní kód, který slouží jako PK, a také atribut název. Tyto dva atributy jsou automaticky vytvořeny u každé entity v MDS. Dále entita účtů obsahuje odkaz na svého předchůdce (tzv. parent-child tabulka). Odkaz na entitu výsledkových skupin, obsahuje i entita PKV. Výsledkové skupiny taktéž definují svého předchůdce a také operátor, který je na skupinu aplikován (+ nebo -)



Obrázek 4-33: Návrh datového modelu MDS pro mapování útvarů

Zdroj: vlastní

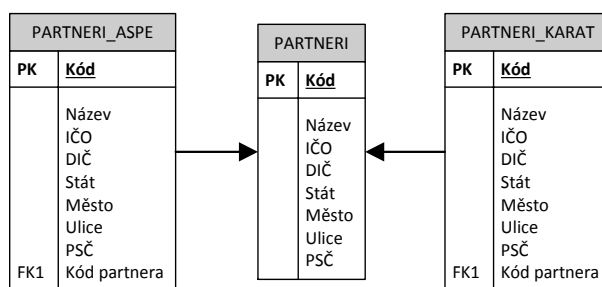
Obrázek 4-33 ukazuje model pro mapování firemních útvarů. Podřízené entity přebírají hierarchickou strukturu ze zdrojových systémů, ale u výstupní entity musí být definována manuálně.



Obrázek 4-34: Návrh datového modelu MDS pro mapování zakázek na stavby

Zdroj: vlastní

Dalšími seznamy, které jsou potřeba mapovat mezi Aspe a KARATem jsou seznamy staveb a stavebních objektů. V tomto případě však výstupní seznam neslouží jako zdroj dimenze, ale pouze jako reference pro vazbu na zakázky z ekonomického programu. Entita staveb a objektů je načtena z Aspe a zakázky z KARATu. Jelikož KARAT neumožňuje podrobnější členění zakázek, bylo nutné pro jejich podrobnější sledování využít jiný analytický rozměr zvaný činnosti. Ty budou v účetnictví reprezentovat stavební objekty.



Obrázek 4-35: Návrh datového modelu MDS pro mapování obchodních partnerů

Zdroj: vlastní

Mapování obchodních partnerů – viz obrázek 4-35, je ze všech uvedených modelů nejsložitější na údržbu. Bude pravděpodobně obsahovat nejvíce záznamů a kvůli velkému počtu atributů je obtížné udržovat výstupní seznam aktuální. Je to však jediná možnost, jak se efektivně vypořádat s redundantními záznamy a mít kontrolu nad tím, jaká data se

do dimenze partnerů dostanou. Z průběžné analýzy stávajících dat bylo zjištěno, že právě číselník partnerů obsahuje velké množství duplicitních avšak ne shodných údajů. Díky této struktuře entit je možné k jednomu záznamu ve výstupním seznamu přiřadit libovolný počet vstupních, čímž bude v DWH zachována unikátnost prvků dimenze a tím pádem i vyšší kvalita výstupů.

Zbývající dimenze pocházející pouze z jednoho systému, je možné naplnit jednoduchým přenosem dat a případnou transformací datových typů.

4.7.2.1 Plnění tabulek faktů

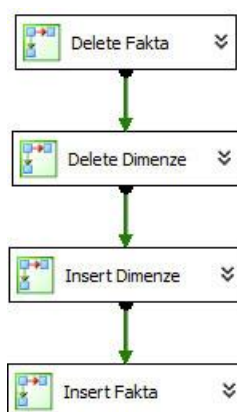
Téměř veškeré fakty definované v datovém modelu pocházejí s Aspe, z ERP KARAT pochází pouze obraty jednotlivých účtů. Při prvotním návrhu byla data čerpána přímými dotazy do zdrojových databází, ale to se ukázalo jako špatná cesta. Při jakékoliv aktualizaci systému bylo nutné dotaz uvnitř datové pumpy změnit, což se ukázalo jako velice neefektivní. S vývojovým týmem Aspe jsme se tedy dohodnuli na jasně stanoveném formátu dat, která budou pomocí dohodnutých databázových pohledů, funkcí nebo procedur prostřednictvím jejich databáze poskytovat. Jedná se tedy o jakési rozhraní mezi Aspe a datovým skladem. Ze strany Aspe bude tedy vždy zajištěn správný formát dat, která se budou do DWH skrz ETL plnit. Databáze KARATu ale žádné podobné rozhraní nenabízí, a proto je v případě této databáze potřeba uložit přímé databázové dotazy přímo do programu datové pumpy. Z této databáze však pochází data pouze pro jednu tabulku faktů a několik sdílených dimenzí, takže případně změny těchto několika málo dotazů by neměly být nijak složité.

4.7.3 SQL Server Integration Services

Pro realizace ETL procesů je v rámci MS SQL Serveru dostupná služba zvaná Integration Services. Pomocí přehledného grafického rozhraní programu MS Business Intelligence Development Studio (BIDS) a velkého počtu transformačních úkonů je možné namodelovat téměř jakýkoliv proces. Tento nástroj jsem při návrhu ETL použil i já.

ETL proces byl navržen tak, aby bylo možné ho jednoduše upravit a plnění DWH probíhalo konzistentně. Proto je celý proces rozdělen do několika dílčích bloků – viz

obrázek 4-36. V prvním kroku probíhá mazání dat z tabulek faktů a poté z tabulek dimenzí. V druhém kroku probíhá plnění tabulek dimenzí a ve třetím pak plnění tabulek faktů. Přesto, že se v datových skladech mazání záznamů běžně nepoužívá kvůli uchovávání historických dat, zde je použití mazání nejvhodnější variantou. Firma nepožaduje sledování změn v datech a ani historizaci dimenzí, proto je rychlejší řešení naplnění DWH znovu, než všechny záznamy kontrolovat na změny a případně je aktualizovat nebo přidat. Dalším důvodem mazání je nekonzistentní výstup dat pro dimenzi potřeb z Aspe. Jelikož tato data nemají jednoznačný stálý identifikátor, jejich unikátní identifikátor je generován vždy při sestavení výstupu z databáze. Proto je aktualizace takových záznamů nemožná a je tedy třeba je vždy přepsat. Na následujícím obrázku je návrh čtyř hlavních skupin operací vytvořených v SSIS. Uvnitř každého z dílčích procesů jsou podrobněji zpracovány transformační operace pro zajištění správného chodu celého ETL.



Obrázek 4-36: Návrh ETL v SSIS

Zdroj: vlastní

Proces transformace byl navržen tak, aby každý ze zmíněných kroků procesu probíhal v izolované transakci. Díky podpoře distribuovaných transakcí napříč různými databázovými servery je možné uzavřít do jedné transakce skupinu dotazů, které jsou zpracovávány na různých serverech. Tato funkcionality je v systému Windows zajišťována službou DTC (Distributed Transaction Coordinator). Službu je však potřeba dodatečně spustit a nakonfigurovat, neboť v základním nastavení systému není spuštěna.

4.8 Modelování datové kostky

Jedním z požadavků na funkcionalitu BI aplikace byla co možná nejrychlejší odezva a získávání výsledků na základě předpočítaných agregovaných hodnot. Tuto funkcionalitu je možné zajistit pomocí OLAP databází.

4.8.1 SQL Server Analysis Services

V rámci SQL Serveru je k dispozici nástroj pro modelování datových kostek zvaný Analysis Services. Pomocí jednoduchého grafického rozhraní je možné definovat uživatelské dimenze s velkým množstvím parametrů a vytvářet velmi komplexní datové kostky.

4.8.2 Dimenze

Dimenze datové kostky vycházejí opět z návrhu. Jak jsou detailně rozpracovány o několik kapitol výše, tak budou naprogramovány i v rámci datové kostky. SSAS umožňuje velmi podrobné nastavení všech prvků dimenze, což se využije především u parent-child dimenze výsledkových skupin. Pro nepravidelné agregace jednotlivých větví je potřeba specifikovat atribut nesoucí označení operátoru agregací (unary operator).

4.8.3 Fakty

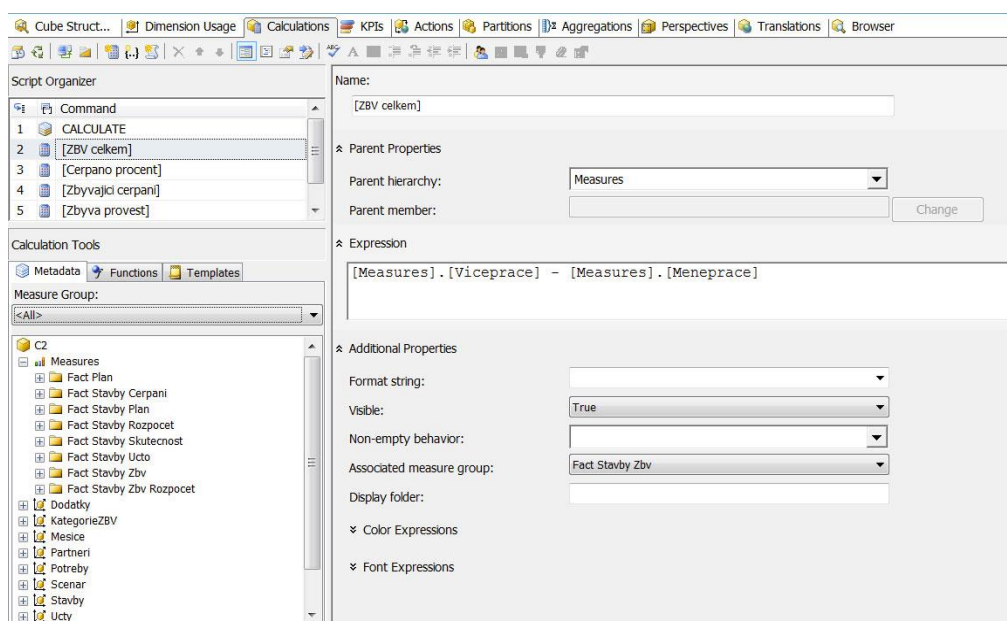
Fakty obsažené v datové kostce opět zcela vyplývají z návrhu datového modelu. Všechny zmíněné fakty jsou buď zcela, nebo alespoň částečně agregovatelné a pro vytváření vyšších agregací je na všechna aplikována funkce sčítání (SUM). Z dostupných fyzických faktů je možné stanovit celou řadu odvozených neboli kalkulovaných faktů. Ty je možné pomocí SSAS velmi jednoduchým způsobem definovat. V následující tabulce jsou uvedeny některé navržené kalkulované fakty.

Tabulka 4-3: Kalkulované fakty

$\text{odbytová cena (OC) stavby} = \sum \text{cena (stavby rozpočet)}$
$\text{skutečné náklady stavby} = \sum \text{obratů z účetnictví za větev nákladů přes dimenzi výsledkových skupin}$
$\text{plánované zbývající náklady stavby} = \sum \text{stavby} - \text{plán za větev nákladů přes dimenzi výsledkových skupin}$
$\begin{aligned} \text{předpokládaný hospodářský výsledek stavby} \\ = \text{OC stavby} - (\text{Skutečné náklady stavby} \\ + \text{Plánované zbývající náklady stavby}) \end{aligned}$
$\text{ZBV celkem} = \text{vícepráce} - \text{méněpráce}$
$\text{zbývající čerpání} = \text{cena aktuální} - \text{cena čerpáno}$
$\text{čerpáno (\%)} = \text{cena čerpáno} / \text{cena aktuální}$

Zdroj: vlastní

Tyto kalkulované fakty je dále možné doplnit o libovolný počet dalších, díky velice flexibilnímu řešení datových kostek. Kalkulace je možné provádět se všemi dostupnými fyzickými fakty. Obrázek 4-37 zachycuje obrazovku programu BIDS, konkrétně prostředí pro definice kalkulovaných faktů uvnitř datové kostky.



Obrázek 4-37: Návrh kalkulovaných faktů v SSAS

Zdroj: vlastní

4.9 Reporty

Poslední částí návrhu BI aplikace je definice výstupů a reportů. Jedná se v podstatě o jedinou část aplikace, se kterou přijde uživatel do kontaktu, neboť tak si to zadávající společnost přeje. Jejím požadavkem při návrhu bylo, aby jediným uživatelským přístupovým bodem do aplikace byl dashboard s reporty.

4.9.1 SQL Server Reporting Services

Pro definici reportů byl zvolen poslední nástroj poskytovaný SQL serverem zvaný Reporting Services. Jedná se opět o aplikaci s příjemným grafickým prostředím, ve kterém mohou programátoři poměrně jednoduše pomocí předpřipravených komponent vytvořit přehledný report z téměř libovolného datového zdroje.

Vytváření reportů zde probíhá v několika krocích. V prvním kroku je potřeba specifikovat datový zdroj, ze kterého budou data do reportu čerpána. Těchto datových zdrojů může být v rámci projektu/reportu použito i více. V tomto případě bude jediným datovým zdrojem OLAP databáze s připravenou datovou kostkou. Dalším krokem je vytvoření tzv. DataSetu neboli kolekce dat, která bude sloužit jako zdroj konkrétních dat pro vybrané komponenty. Díky nativní podpoře datového zdroje typu MSOLAP a přehlednému vizuálnímu návrháři MDX dotazů s podporou parametrů je vytvoření DataSetu otázkou několika málo kliknutí. V posledním kroku při vytváření reportů je potřeba vytvořený DataSet propojit s vybranou komponentou (tabulka, matice, budík, mapa atd.) a vše graficky vyladit. Poté je již možné vytvořený report odeslat na reportingový server a nasadit jej do ostrého provozu a zpřístupnit uživatelům.

Pomocí připravených dat je možné definovat téměř jakékoliv interaktivní reporty. Jedním z požadavků na výstupy z aplikace byl report obsahující rekapitulaci čerpání za jednotlivé stavby. Obsahuje název stavby, původní plánovanou cenu, čerpanou částku, procento čerpání a semafor indikující stav čerpání. Dalším požadavkem je zeleně zobrazit stavby, které doposud nepřesáhly čerpanou částku a červeně ty, které ji přesáhly. Obrázek níže ukazuje, jak takový vzorový report rekapitulace čerpání může vypadat.

Rekapitulace čerpání staveb

Stavba	Cena původní	Částka čerpáno	Čerpáno [%]	Zbývá k čerpání	Stav
0135 ZVS - Dálnice D1 stavba 0135 Kroměříž východ - Říkovice	4 993 486 798Kč	4 510 866 620Kč	90,34%	482 620 178Kč	●
02/2007 - I/23 TELČ PRŮTAH	26 507 229Kč	12 176 810Kč	45,94%	14 330 420Kč	●
0219-08-3 - I/38 DLOUHÁ BRTNICE - MOST EV.Č.38-087	7 797 541Kč	2 511 295Kč	32,21%	5 286 245Kč	●
0305 II. etapa - Dálnice D3-0305-II II. etapa PDPS	548 315 255Kč	500 017 142Kč	91,19%	48 298 113Kč	●
03-233-2-000 - R4 křižovatka I/20 Nová Hospoda	1 062 345 389Kč	1 155 434 362Kč	108,76%	0Kč	●
03-720-00 - I/24 Třeboň, most ev. č. 24-008.3	59 902 073Kč	59 902 073Kč	100,00%	0Kč	●
04-011-2-000 - 3272231006 D8 stavba 0805 část B - most	441 677 364Kč	470 217 730Kč	106,46%	0Kč	●
04-012-2-000 - D8 stavba 0805 část C - most Oparno	423 113 912Kč	336 177 628Kč	79,45%	86 936 283Kč	●
04-013-2-000 - D8 stavba 0805 část D - most Dobkovičky	600 777 130Kč	712 410 066Kč	118,58%	0Kč	●

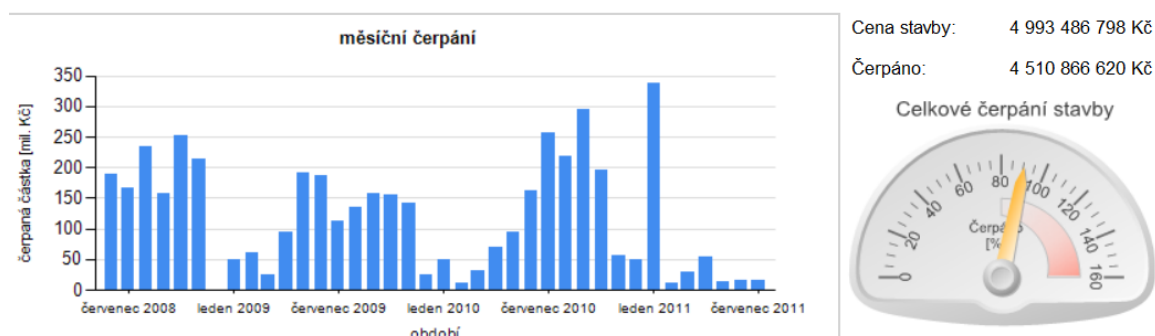
Obrázek 4-38: Návrh reportu rekapitulace čerpání

Zdroj: vlastní

Tento report je dále možné rozšířit například o funkci pro zobrazení detailnějších informací. Po kliknutí na název stavby bude uživatel přesměrován na další report, na kterém jsou zobrazeny další podobnější informace o čerpání zvolené stavby. Návrh takového reportu se nachází na následujícím obrázku. Report obsahuje název zvolené stavby, graf znázorňující čerpání v jednotlivých měsících, celkovou cenu stavby, celkovou čerpanou částku a budíkem znázorněné procento čerpané částky.

Detail čerpání stavby

0135 ZVS - Dálnice D1 stavba 0135 Kroměříž východ - Říkovice



Obrázek 4-39: Návrh reportu detailu čerpání stavby

Zdroj: vlastní

5 Zhodnocení přínosu

Přes velké množství komerčně nabízených BI řešení neexistuje takové, které by zcela vyhovovalo potřebám společnosti BREX. Za hlavní přínos této práce tak lze považovat vytvoření návrhu BI aplikace, která požadované splňuje.

Nosnou částí celé aplikace je datový model. Ten je díky použité Kimballově metodě multidimenzionálního modelování velmi pružný a řešení je možné kdykoliv poměrně jednoduše rozšířit a doplnit o další funkcionalitu.

Zdárné dokončení tohoto projektu by firmě mělo umožnit mnohem lepší sledování nákladů všech nově realizovaných stavebních zakázek a rychle získat ucelený přehled o fungování firmy v porovnání oproti plánu. Díky rychlejším a ucelenějším přehledům ohledně hospodaření podniku by se firmě mohla také zvýšit konkurenceschopnost mezi stavebními firmami.

5.1 Ekonomické zhodnocení

Úkolem této podkapitoly je, zhodnotit navržené řešení z ekonomické stránky. Náklady na projekt je možné rozdělit do dvou kategorií. Pořizovací náklady a provozní.

5.1.1 Pořizovací náklady

Firma při realizaci toho projektu zvolila cestu vlastního vývoje. Přibližné vynaložené dosavadní pořizovací náklady jsou srovnány s náklady na realizaci pomocí externí firmy a cenami hotových řešení.

Veškerá doba, kterou jsem strávil studiem informací, analýzou požadavků, návrhem konkrétních řešení a jejich konzultací a optimalizací je přibližně 500 hodin. Při průměrné výši nástupního platu zaměstnanců s nižším vysokoškolským vzděláním cca 15 000 Kč

hrubého/měsíc (z nejmenovaného zdroje), jsou celkové měsíční náklady zaměstnavatele na takového zaměstnance rovných 20 100 Kč. Náklady na mou osobu se tedy při uvedené časové zátěži a mzdě rovnají 62 800 Kč. Tato částka ovšem není konečná. Práce na projektu se neobešla bez nutných konzultací s několika vrcholovými pracovníky a ostatními zaměstnanci. Časová zátěž manažerů byla stanovena na 30 hodin a ostatních řadových zaměstnanců na 50 hodin. Dále je k této částce potřeba přičíst ostatní režijní náklady na zaměstnance v celkové výši cca 12 000 Kč. Sečtením všech zmíněných složek vyjde částka rovných 100 000 Kč, která je rovna dosavadním nákladům vynaložených firmou na realizaci tohoto projektu, kde se v současné době již projekt nachází ve fázi testování částečně realizovaného řešení. Rekapitulace jednotlivých složek pořizovacích nákladů je uvedena v následující tabulce.

Tabulka 5-1: Rozpis pořizovacích nákladů

Náklad	Popis výpočtu	Celková cena [Kč]
Má osoba	Náklady firmy vynaložené za mojí osobu za 500 hodin práce při měsíční mzdě 15 000 Kč hrubého.	62 800
Konzultace s manažery	Náklady za práci manažerů, kteří se na řešení projektu podíleli. (při uvažování průměrného platu 50 000 Kč a časové zátěži 30 hodin)	12 600
Konzultace s ostatními pracovníky	Náklady za práci ostatních zaměstnanců, kteří se na řešení projektu podíleli. (při uvažování průměrného platu 30 000 Kč a časové zátěži 50 hodin)	12 600
Ostatní režijní náklady	Ostatní režijní náklady – např. administrativa, nájem, energie atd.	12 000
CELKEM:		100 000

Zdroj: vlastní

Z informací dostupných na webových stránkách několika programátorských společností bylo zjištěno, že každá započatá hodina nespecifikovaného programátora se pohybuje okolo 1 000 Kč bez DPH, spíše výše. Pokud touto cenou vynásobíme celkový čas strávený analýzou, návrhem a realizací (500 hodin), dostaneme cenu 500 000 Kč. Dá se

předpokládat, že programátor nebo tým programátorů, specializovaných na tuto problematiku, by stejný projekt vyřešil rychleji. Pokud vezmeme v úvahu, že by specializovaná firma zvládla stejný úkol za poloviční čas, dostaneme částku 250 000 Kč.

Ceny dostupných hotových BI řešení jsou velmi různé. Od desítek tisíc (za malá řešení typu nadstavba nad konkrétním ekonomickým softwarem³⁵) až po stovky tisíc (za velká specializovaná řešení pro velké společnosti³⁶). Jelikož aplikace tohoto typu je zcela specifická a nemá obdoby, s cenami hotových řešení ji nelze příliš srovnávat. Navíc hotová řešení jsou většinou připravena k okamžité implementaci a tento projekt se v této fázi ještě nenachází. Cenu samostatného datového modelu se zjistit nepodařilo.

Porovnáme-li přibližné náklady vynaložené firmou za vlastní vývoj s možnými náklady za vývoj specializovanou firmou zjistíme, že specialisté by byli minimálně 2,5 krát dražší, než dosavadní cena projektu. Z toho lze vyvodit, že zvolená cesta vývoje je z mého pohledu výrazně levnější. Klady a zápory zvolené cesty vývoje oproti jiným možnostem mohou být předmětem dalších rozsáhlých diskuzí, které však nejsou předmětem této práce.

5.1.2 Provozní náklady

Teoreticky lze říci, že provozní náklady budou v závislosti na těch pořizovacích neměnné. Existuje však mnoho proměnných, na kterých toho tvrzení závisí. Během ekonomického zhodnocení tedy bude uvažováno, že jak řešení navržené externí firmou tak i již hotové nabízené řešení bude stejné, jako je navrženo v této práci.

Provozní náklady takovéto aplikace se skládají z pravidelných údržeb serveru, na kterém je aplikace provozována a v tomto konkrétním případě především z úkonů spojených se správou kmenových dat. Pro správnou funkčnost a zajištění kvality výstupů musí být tato data často kontrolována, doplňována a aktualizována. Při týdenních intervalech, časové zátěži cca 3h a hodinových mzdových nákladech na zaměstnance 168 Kč budou tyto

³⁵ Ceník řešení POHODA Business Intelligence 2014. *POHODA: Ekonomický systém* [online]. 2014 [cit. 2014-04-24]. Dostupné z: <http://www.stormware.cz/pohoda/business-intelligence/cenik/>

³⁶ Oracle snižuje ceny softwaru pro Business Intelligence. *Channel World: Zpravodajství pro prodejní kanály IT / CE* [online]. 2013 [cit. 2014-04-24]. Dostupné z: <http://channelworld.cz/zpravy/oracle-snizuje-ceny-softwaru-pro-business-intelligence-9119>

náklady přibližně 2000 Kč/měs. Náklady na údržbu serveru se špatně odhadují, ale při využití vlastních zdrojů byly stanoveny na 500 Kč/měs. Celkové provozní náklady této BI aplikace se tedy rovnají přibližně 2500 Kč/měs.

5.1.3 Návratnost investice

V současné době provádí firma controlling probíhajících stavebních zakázek ručně, přičemž tato činnost zaměstná hlavního ekonoma společnosti každý měsíc přibližně na celé dva dny. Tento čas je závislý na počtu kontrolovaných zakázek, který se běžně pohybuje mezi 15 až 20. Při uvažovaném platu 40 000 Kč hrubého/měsíc vyjde tato práce firmu každý měsíc přibližně na 5 600 Kč. Návratnost doposud vynaložené investice se tedy v tomto případě rovná 22 měsícům. Pokud do výpočtu zahrneme nově vzniklé náklady na správu kmenových dat, návratnost se zvýší na 39 měsíců.

Návratnost investice do vybudování podnikového DWH a nasazení BI nástrojů však může být i mnohem rychlejší. To především díky nově získaným úhlům pohledu, on-line přístupným datům, rychlým reportům a interaktivním dashboardům. Díky mnohem lepší možnosti sledování nákladů při controllingu, který je díky této aplikaci dostupný, může firma, dle vyjádření managementu, během realizace zakázek ušetřit desítky až stovky tisíc.

Závěr

Hlavním cílem bakalářské práce bylo vytvoření návrhu BI aplikace pro stavební společnost BREX. Hlavní cíl byl dále rozdělen do několika dílčích cílů, které celou problematiku dělilo na několik menších ucelených částí. Prvním dílčím cílem bylo vytvoření návrhu modelu datového skladu. Počáteční analýza požadavků byla velmi zdoluhavá, ale po získání všech potřebných odpovědí bylo dosaženo požadovaného cíle a datový model se podařilo vytvořit. Ten byl vytvořen na základě rozšířené BUS matice, která je využívána v rámci Kimballovy metodologie multidimenzionálního modelování.

Druhým dílčím cílem bylo sestavení návrhu ETL procesu, který bude zajišťovat plnění DWH daty z databází ASW Aspe a ERP KARAT. Proces byl navržen pouze metodicky, jak by přenos dat měl probíhat. Konkrétní aplikaci je potřeba zpracovat na základě přesných databázových výstupů. V rámci návrhu ETL byla navržena metoda mapování pro sjednocování a párování stejných záznamů pocházejících z různých systémů. Při použití této metody se do DWH dostávají pouze zkontrolovaná a očištěná data.

Třetím dílčím cílem bylo navržení OLAP databáze. Jelikož jsou modely dimenzí a faktů identické s těmi v návrhu datového skladu, návrh definuje pouze několik kalkulovaných faktů, které je možné v prostředí SSAS definovat a dle libosti dále rozšiřovat.

Posledním dílčím cílem bylo navržení reportů. Jelikož jsou možnosti reportů takřka neomezené a jediným limitem jsou dostupná data (a je tedy možné vytvořit konkrétní reporty zcela dle požadavků zákazníka), navrhl jsem pouze jeden vzorový report pro nastínění funkcionality a možností SSRS.

V současné době se projekt nenachází pouze ve fázi návrhu, ale již dochází k testování reálně zpracovaných komponent. Datový model je již fyzicky zpracován pomocí relační databáze MS SQL Server 2008 R2 a vytvořený ETL proces dle navrženého modelu v prostředí SSIS do DWH přenáší určená data z obou zdrojových systémů. OLAP databáze

je již také realizována a dochází pouze k drobným úpravám. Reporty jsou stále ve stádiu návrhu a budou zpracovány až po dokončení datové kostky.

Jelikož je projekt řešen přímo na míru dané společnosti, jeho nasazení v jiných firmách by mohlo být problematické kvůli jiným zavedeným podnikovým procesům. Řešení je postaveno tak, aby umožňovalo analýzu co největšího množství informací z Aspe. Napojení jiného výrobního ASW do stávající datové struktury pravděpodobně není možné, avšak ekonomická část modelu je navržena flexibilně a napojení jakéhokoliv ekonomického systému by obnášelo pouze úpravu malé části ETL.

Z ekonomického pohledu se zvolená cesta vývoje firmě vyplatila, neboť vynaložené náklady představují pouze malou část toho, co by projekt mohl stát u specializované firmy. Zakoupení hotového řešení by mohlo být levnější, ale nebylo by s ním možné provádět tak specifické analýzy jako při vytvoření řešení přesně na míru. Návratnost investice byla vypočtena na 22, resp. 39 měsíců, ale je očekávána spíše dříve.

Díky flexibilnímu datovému modelu je aplikace snadno rozšiřitelná. Za zvážení by stálo doplnění více ekonomických ukazatelů, definování KPI nebo rozšíření o možnost detailnější analýzy účetních výkazů.

Seznam použité literatury

Citace

1KEYDATA. Bill Inmon vs. Ralph Kimball. *1KeyData* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.1keydata.com/datawarehousing/inmon-kimball.html>

BREX: Ekonomické informace. B R E X, spol s. r.o. *BREX: stavební společnost* [online]. 2013 [cit. 2014-02-09]. Dostupné z: <http://www.brex.cz/cs/spolecnost/ekonomicke-informace/>

BREX: Profil společnosti. B R E X, spol s. r.o. *BREX: stavební společnost* [online]. 2013 [cit. 2014-02-09]. Dostupné z: <http://www.brex.cz/cs/spolecnost/profil-spolecnosti/>

DANEL, Roman. *Dolování dat* [online]. 2010, 3 s. [cit. 2014-05-05]. Dostupné z: http://homel.vsb.cz/~dan11/is_skripta/IS%202010%20-%20Danel%20-%20Dolovani%20dat.pdf

FROFINIT. *Dimenzionální modelování*. 2013. Dostupné z: http://www.profinet.eu/fileadmin/Content/profinet.eu/Academy/MI-DSP/08_Dimenzionalni_modelovani.pdf

HORÁK, Jiří a Bronislava HORÁKOVÁ. *Datové sklady a využití datové struktury typu hvězda pro prostorová data*. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava, 2007. Dostupné z: http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2007/sbornik/Referaty/Sekce3/hvezdaF4.pdf

CHEE, Timothy. *BUSINESS INTELLIGENCE SYSTEMS: STATE-OF-THE-ART REVIEW AND CONTEMPORARY APPLICATIONS*. In: Symposium on Progress in Information & Communication Technology 2009. Malaysia: xxx, 2009, s. 96. Dostupné z: http://spict.utar.edu.my/SPICT-09CD/contents/pdf/SPICT09_A-5_1.pdf

IPOS-SOFT. *IPOS-SOFT: Komplexní informační a řídicí systémy pro investiční výstavbu* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.ipossoft.cz/>

KARAT SOFTWARE A. S. *Karat: informační systém* [online]. 2014 [cit. 2014-05-03]. Dostupné z: <http://www.karatsoftware.cz/>

KIMBALL, Ralph. Drilling Down, Up, and Across: Understanding the vocabulary of navigating dimensions. *KIMBALL GROUP* [online]. 1996 [cit. 2014-05-05]. Dostupné z: <http://www.kimballgroup.com/1996/03/01/drilling-down-up-and-across/>

KIMBALL, Ralph a Margy ROSS. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2nd ed. New York: Wiley, c2002. ISBN 0471200247.

KIMBALL, Ralph a Margy ROSS. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Third edition. Indianapolis: John Wiley & Sons, Inc., 2013. ISBN 11-185-3080-2.

LABERGE, Robert. *Datové sklady: agilní metody a business intelligence*. 1. vyd. Brno: Computer Press, 2012, 350 s. ISBN 978-80-251-3729-1.

LUHN, H. P. *A Business Intelligence System*. IBM Journal of Research and Development. 1958, vol. 2, issue 4. DOI: 10.1147/rd.24.0314. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5392644>

MELICHAR, Jan. *Implementace Business Intelligence ve stavebnictví*. Praha, 2008. Dostupné z: http://www.vse.cz/vskp/show_file.php?soubor_id=23910. Diplomová práce. Vysoká škola ekonomická v Praze.

MICROSOFT. *Microsoft SQL Server* [online]. 2012 [cit. 2014-04-18]. Dostupné z: <http://www.microsoft.com/sqlserver/cs/cz/default.aspx>

MICROSOFT. Partition Storage Modes and Processing. MICROSOFT. *Microsoft Developer Network* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://msdn.microsoft.com/en-us/library/ms174915.aspx>

POUR, Jan, Miloš MARYŠKA a Ota NOVOTNÝ. *Business intelligence v podnikové praxi*. 1. vyd. Praha: Professional Publishing, 2012. ISBN 978-80-7431-065-2.

REEVES, Laura L. *A Manager's Guide to Data Warehousing*. Indianapolis, IN: Wiley Pub., c2009. ISBN 0470176385.

TERADATA. Teradata Database: Teradata Database 15.0. TERADATA [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.teradata.com/products-and-services/Teradata-Database>

VALBEK. Co je Aspe. *Aspe: Software pro řízení staveb* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.aspe.cz/cs/produkty/aspe/co-je-aspe>

WITHEE, Ken. *Microsoft business intelligence for dummies*. Hoboken, NJ: Wiley Pub., c2010, xviii, 405 p. ISBN 978-0-470-52693-4. Dostupné z: <http://it-ebooks.info/go.php?id=1388-1399142562-67da65983c27f00ebe250633202c0f70>

ZÁDOVÁ, Vladimíra. *Specifika postavení a návrhu datových skladů v rámci IS/ICT*. Liberec, 2006. Disertační. Technická univerzita v Liberci

Bibliografie

AXERO SOLUTIONS LLC. *BIDN* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://bidn.com/>

KIMBALL GROUP. *KIMBALL GROUP* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.kimballgroup.com/>

LACKO, Luboslav. *Business Intelligence v SQL Serveru 2005: reportovací, analytické a další datové služby*. Vyd. 1. Brno: Computer Press, 2006, 391 s. ISBN 80-251-1110-5.

LACKO, Luboslav. *Business Intelligence v SQL Serveru 2008: reportovací, analytické a další datové služby*. Vyd. 1. Brno: Computer Press, 2009, 456 s. ISBN 978-80-251-2887-9.

MICROSOFT. *Microsoft Developer Network* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://msdn.microsoft.com/>


SIMPLE TALK PUBLISHING. *SQL Server Central* [online]. 2014 [cit. 2014-05-05]. Dostupné z: <http://www.sqlservercentral.com/>

Seznam příloh

Příloha A – Schémata návrhu tabulek dimenzí databáze datového skladu


Příloha B – Schémata návrhu tabulek faktů a jednotlivých DM datového skladu

Příloha A – Schémata návrhu tabulek dimenzí databáze datového skladu

dim_mesice			
Column Name	Data Type	Allow Nulls	
 id_dim_mesice	int	<input type="checkbox"/>	
kalendarni_rok	smallint	<input type="checkbox"/>	
kalendarni_rok_nazev	varchar(10)	<input checked="" type="checkbox"/>	
ucetni_rok	smallint	<input type="checkbox"/>	
ucetni_rok_nazev	varchar(10)	<input checked="" type="checkbox"/>	
kalendarni_ctvtleti	tinyint	<input type="checkbox"/>	
kalendarni_ctvtleti_nazev	varchar(20)	<input checked="" type="checkbox"/>	
ucetni_ctvtleti	tinyint	<input checked="" type="checkbox"/>	
ucetni_ctvtleti_nazev	varchar(20)	<input checked="" type="checkbox"/>	
mesic	tinyint	<input type="checkbox"/>	
mesic_nazev	varchar(20)	<input checked="" type="checkbox"/>	
		<input type="checkbox"/>	


Obrázek A-1: Tabulka dimenze času (měsíců)

Zdroj: vlastní

dim_utvary			
Column Name	Data Type	Allow Nulls	
 id_dim_utvary	int	<input type="checkbox"/>	
k_predch	int	<input checked="" type="checkbox"/>	
znacka	varchar(10)	<input checked="" type="checkbox"/>	
nazev	varchar(128)	<input checked="" type="checkbox"/>	
mesto	varchar(100)	<input checked="" type="checkbox"/>	
		<input type="checkbox"/>	

Obrázek A-2: Tabulka dimenze útvarů

Zdroj: vlastní

dim_dodatky			
Column Name	Data Type	Allow Nulls	
 id_dim_dodatky	int	<input type="checkbox"/>	
znacka	char(10)	<input checked="" type="checkbox"/>	
nazev	varchar(60)	<input checked="" type="checkbox"/>	
vypracoval	varchar(60)	<input checked="" type="checkbox"/>	
datum_vypracovani	date	<input checked="" type="checkbox"/>	
		<input type="checkbox"/>	

Obrázek A-3: Tabulka dimenze dodatků

Zdroj: vlastní



Obrázek A-4: Tabulka dimenze staveb

Zdroj: vlastní

dim_potreby		
Column Name	Data Type	Allow Nulls
id_dim_potreby	int	<input type="checkbox"/>
skupina_znacka	varchar(50)	<input checked="" type="checkbox"/>
skupina_nazev	varchar(60)	<input checked="" type="checkbox"/>
potreba_znacka	varchar(50)	<input checked="" type="checkbox"/>
potreba_nazev	varchar(160)	<input checked="" type="checkbox"/>
mj_znacka	varchar(50)	<input checked="" type="checkbox"/>
mj_nazev	varchar(50)	<input checked="" type="checkbox"/>
jc	money	<input checked="" type="checkbox"/>


Obrázek A-5: Tabulka dimenze potřeb

Zdroj: vlastní

dim_kategorie_zbv		
Column Name	Data Type	Allow Nulls
id_dim_kategorie_zbv	int	<input type="checkbox"/>
znacka	varchar(15)	<input checked="" type="checkbox"/>
nazev	varchar(200)	<input checked="" type="checkbox"/>


Obrázek A-6: Tabulka dimenze kategorií ZBV

Zdroj: vlastní

dim_partneri		
Column Name	Data Type	Allow Nulls
 id_dim_partneri	int	<input type="checkbox"/>
nazev	varchar(255)	<input checked="" type="checkbox"/>
ico	varchar(15)	<input checked="" type="checkbox"/>
dic	varchar(20)	<input checked="" type="checkbox"/>
mesto	varchar(255)	<input checked="" type="checkbox"/>
ulice	varchar(60)	<input checked="" type="checkbox"/>
psc	varchar(10)	<input checked="" type="checkbox"/>
stat	varchar(3)	<input checked="" type="checkbox"/>
forma	varchar(50)	<input checked="" type="checkbox"/>
druh	varchar(50)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>


Obrázek A-7: Tabulka dimenze partnerů

Zdroj: vlastní

dim_scenar		
Column Name	Data Type	Allow Nulls
 id_dim_scenar	int	<input type="checkbox"/>
nazev	varchar(50)	<input checked="" type="checkbox"/>
verze	int	<input type="checkbox"/>
datum_schvaleni	date	<input checked="" type="checkbox"/>
		<input type="checkbox"/>


Obrázek A-8: Tabulka dimenze scénářů

Zdroj: vlastní

dim_vysl_skupiny		
Column Name	Data Type	Allow Nulls
 id_dim_vysl_skupiny	int	<input type="checkbox"/>
k_predch	int	<input checked="" type="checkbox"/>
znacka	varchar(15)	<input checked="" type="checkbox"/>
nazev	varchar(100)	<input checked="" type="checkbox"/>
operator	char(1)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Obrázek A-9: Tabulka dimenze výsledkových skupin

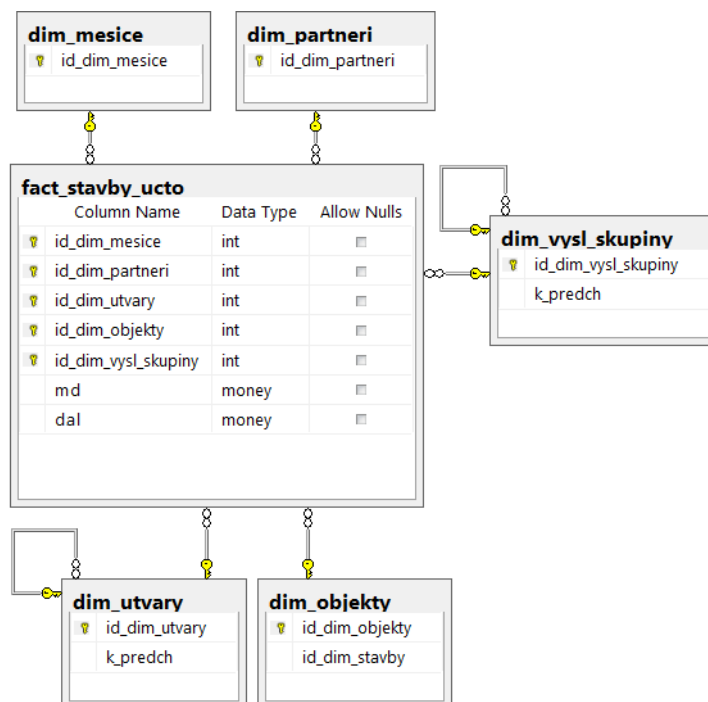
Zdroj: vlastní

dim_zjistovaci_protokoly		
Column Name	Data Type	Allow Nulls
 id_dim_zjistovaci_protokoly	int	<input type="checkbox"/>
znacka	varchar(15)	<input checked="" type="checkbox"/>
datum_vystaveni	date	<input type="checkbox"/>
id_investor	int	<input checked="" type="checkbox"/>
id_dodavatel	int	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Obrázek A-10: Tabulka dimenze zjišťovacích protokolů

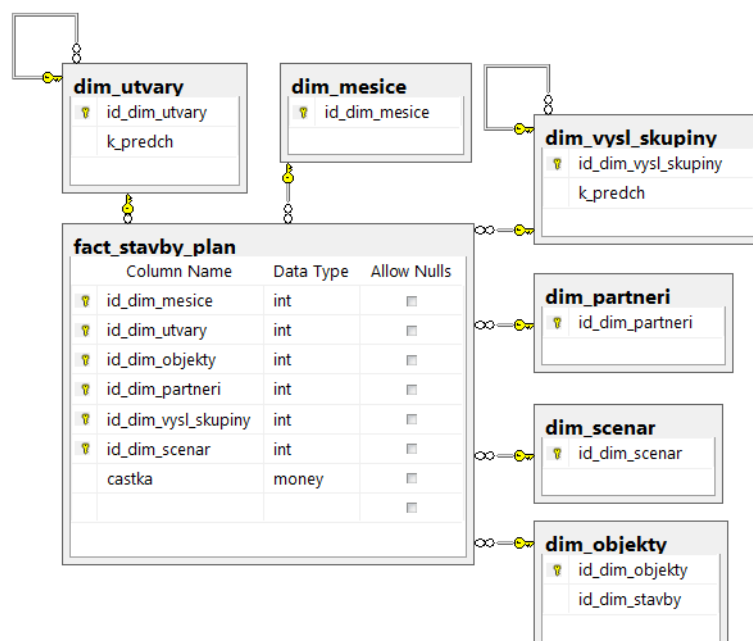
Zdroj: vlastní

Příloha B – Schémata návrhu tabulek faktů a jednotlivých DM datového skladu



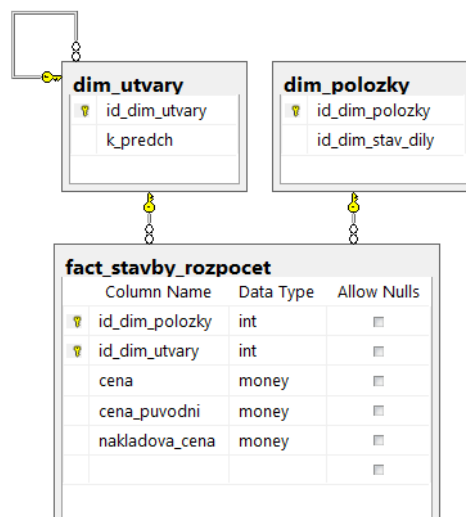
Obrázek B-1: Tabulka faktů pro DM „Obraty staveb v účetnictví“

Zdroj: vlastní



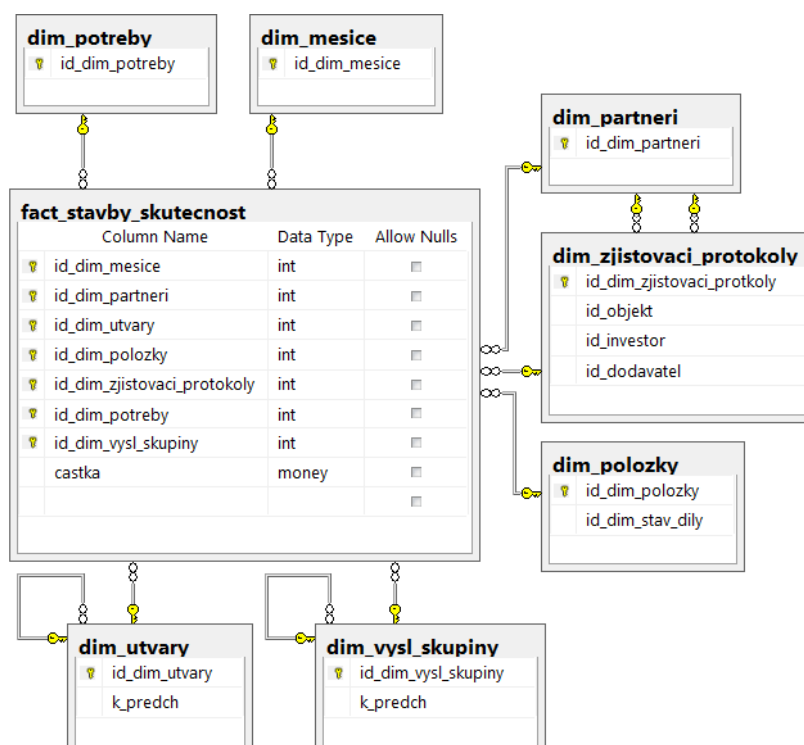
Obrázek B-2: Tabulka faktů pro DM „Plán stavebních zakázek“

Zdroj: vlastní



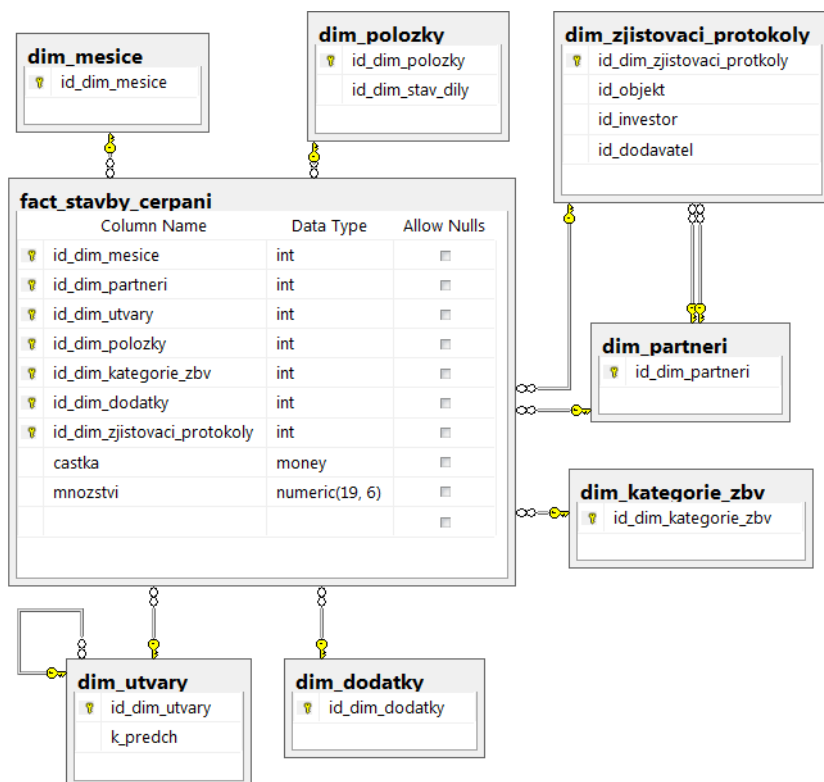
Obrázek B-3: Tabulka faktů pro DM „Rozpočet staveb“

Zdroj: vlastní



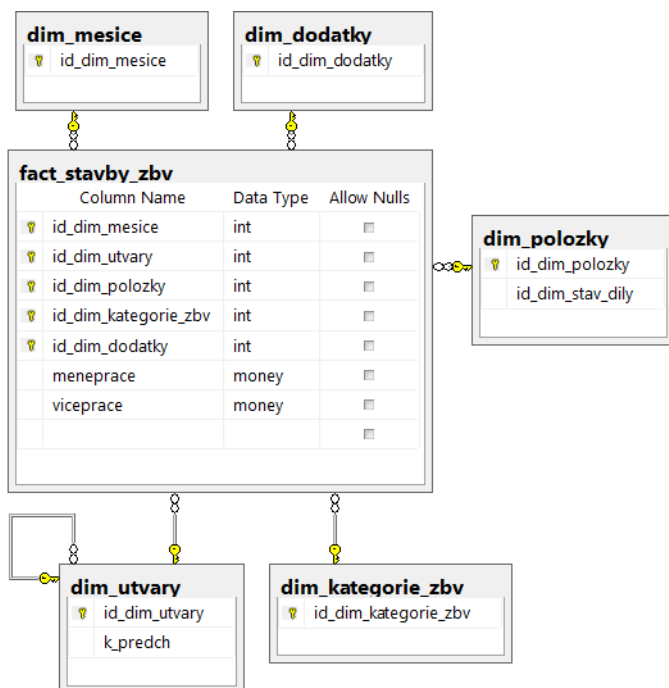
Obrázek B-4: Tabulka faktů pro DM „Skutečná fakturace staveb v Aspe“

Zdroj: vlastní



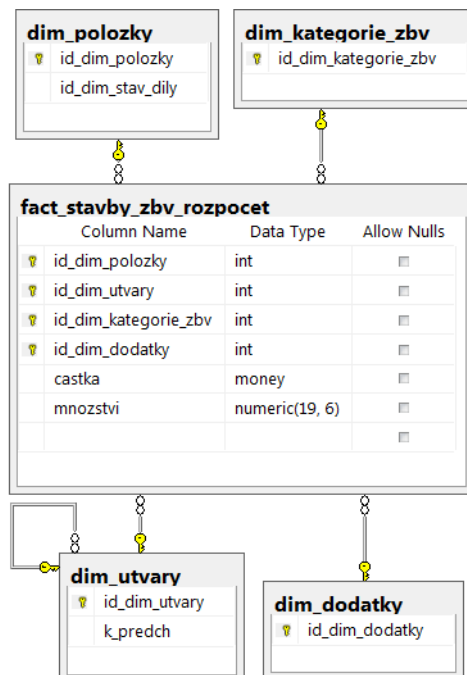
Obrázek B-5: Tabulka faktů pro DM „Čerpání staveb“

Zdroj: vlastní



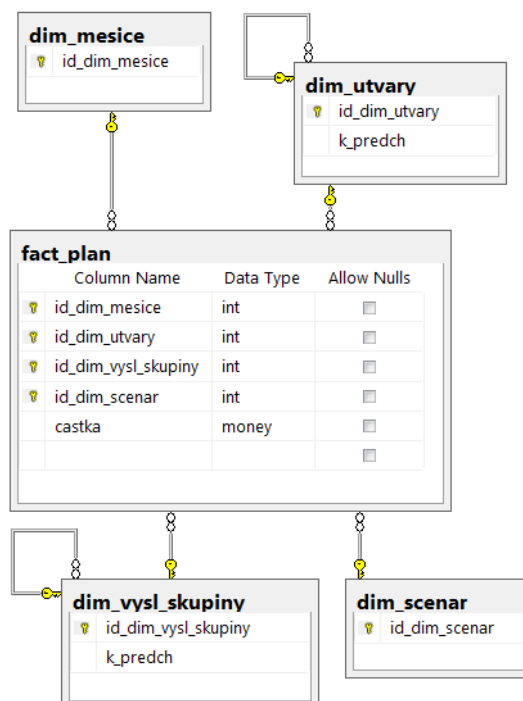
Obrázek B-6: Tabulka faktů pro DM „Změny během výstavby“

Zdroj: vlastní



Obrázek B-7: Tabulka faktů pro DM „Rozpočet ZBV“

Zdroj: vlastní



Obrázek B-8: Tabulka faktů pro DM „Plán firmy“

Zdroj: vlastní